State University of New York at Buffalo
Department of Computer Science

THE AUTOREGRESSIVE METHOD:

A METHOD OF APPROXIMATING AND ESTIMATING POSITIVE FUNCTIONS

by

Jean-Pierre Carmichael*

Statistical Science Division
State University of New York at Buffalo

TECHNICAL REPORT NO. 45
August 1976

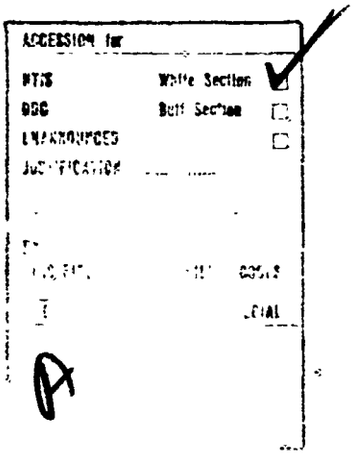
* Research supported in part by the Office of Naval Research.

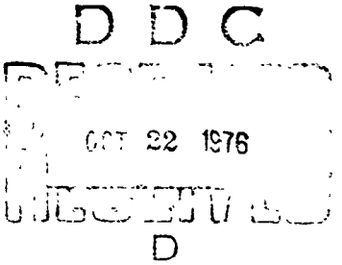
ABSTRACT

We consider the Fourier transform of a positive function $f(\cdot)$ (or its sample Fourier transform) as a possibly complex covariance function of a hypothetical stationary complex-valued time series. We model this time series by an autoregressive process of order p whose spectral density approximates (or estimates) the function $f(\cdot)$.

We show the equivalence of this interpretation with the theory of orthogonal polynomials on the unit circle; we study the consistency of the autoregressive estimator as p increases with the sample size.

We also make an exploratory investigation of this new method as a density estimation method following three approaches: the direct approach, the hazard approach and the sparsity approach.





SUMMARY

— What has been done:

In Part I, we look at the autoregressive method from the practical standpoint. Our aim is twofold:

— to find out how the autoregressive method behaves in many different situations, when the true answer is either known or unknown;

— to compare the autoregressive method to other methods of curve estimation currently used: the kernel method, the spline method, the orthogonal series method and a quantile expansion method.

It is properly impossible to summarize this kind of work. One can form one's opinion only by reading the text and looking at the pictures we produced. Our opinion is that the method is very versatile; it always yields a positive function; it is very easy to use and it works well!

In Part II, we look at the autoregressive method from the theoretical standpoint.

In Chapter 3, we unify the time series interpretation and the orthogonal polynomial interpretation of the autoregressive method. From the time series point of view, our treatment is slightly more general than the current practice in that our autoregressive coefficients are complex numbers.

Chapters 4 and 5 are parallel.

In Chapter 4, we study the different modes of convergence of the autoregressive method, that is the autoregressive method as an approximation method.

The weakest result is Theorem 4.1:

If

$$(1) \qquad \begin{cases} 0 < \int_{-\pi}^{\pi} f(x)\, dx < \infty \\ 0 < \int_{-\pi}^{\pi} \frac{1}{f(x)}\, dx < \infty \\ -\infty < \int_{-\pi}^{\pi} \log f(x)\, dx \end{cases} ,$$

then

$$\lim_{p \to \infty} \int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \cdot f(x)\, dx = 0$$

and

$$\lim_{p \to \infty} \int_{-\pi}^{\pi} \frac{\left| f(x) - f_p(x) \right|}{f_p(x)}\, dx = 0 \quad .$$

The strongest result is Theorem 4.3:

If

$$(2)\qquad \begin{cases} \text{condition (1) holds} \\ 0 < m \le f(x) \le M \quad , & \text{a.e. in } [-\pi,\pi] \\ f(\cdot) = g(\cdot) \quad , & \text{a.e. in } [-\pi,\pi] \\ g(\cdot) \in \mathrm{Lip}(\alpha,2) \quad , & \alpha > \frac{1}{2} \quad , \end{cases}$$

then

$$\lim_{p\to\infty} f_p(x) = \frac{1}{2\pi} \cdot \frac{1}{\left|\pi(e^{ix})\right|^2} \quad , \text{ uniformly} \ .$$

In Chapter 5, we prove the consistency of the autoregressive method as in Theorem 5.9:

If (2) holds and $\lim_{n\to\infty} \frac{p^3}{n} = 0$, then

$$\lim_{n\to\infty} \left| \hat{f}_p(x) - \frac{1}{2\pi} \cdot \frac{1}{\left|\pi(e^{ix})\right|^2} \right| = 0 \quad \text{in probability uniformly.}$$

Finally in Chapter 6, we apply these different results to the problem of density estimation.

— What has to be done:

In the theory:

We have to find the asymptotic distribution of the autoregressive estimator when we allow the order p to increase with the sample size at a given rate. Berk (1974) has worked in this direction. We have weakened some of his assumptions for the consistency of the autoregressive coefficients (see section 5.2), but we haven't touched the other

problem.

Then will come the problem of finding global confidence bounds. We also need a criterion to choose the order p.

In the applications:

The most important task here is probably to justify rigorously the use of the autoregressive method in those cases where $F(\cdot)$ is unbounded (see section 6.2).

We have not touched the problem of estimating the intensity function of a counting process.

TABLE OF CONTENTS

INTRODUCTION

In statistical analysis we are often interested in estimating curves, probably because of the heavy emphasis on visual training in our cultures. The curves are not always drawn but they could be as, for example, in regression analysis: now people more routinely take a look at residual plots to judge the fit to the data provided by the regression line (e.g., Tukey (1970), Feder (1974)).

Actually, there is increased interest in drawing the curves, thanks to computer graphics. A few people have very imaginatively proposed and developed new ways of visualizing the data providing statistical scientists with new means to gain insight into the data and to convey these insights to their clients (e.g., Andrews (1972), Chernoff (1973), Cleveland and Kleiner (1975)).

After a long domination of parametric techniques, we can now let the data speak for itself. That is what the so-called non-parametric techniques are attempting to do. There are already several non-parametric estimators of curves like probability density functions, hazard functions, intensity functions, ...

A universal requirement seems to be the smoothness of the estimators. Indeed, smoothness allows easy integration and differentiation, when required. Estimators should also belong to the class of functions they are trying to estimate. Finally the methods of estimation should be easy to use.

We review briefly the general methods that have appeared in the literature; then we expose the new autoregressive method.

But first a word on the notation. We will stick to the following conventions:

— a function $g(\cdot)$ is approximated by a function $g_m(\cdot)$, where m denotes the order of approximation;

— an approximator $g_m(\cdot)$ is estimated by a function $\hat{g}_m(\cdot)$;

— a function $g(\cdot)$ is estimated directly from a sample by a function $g_n(\cdot)$, where n denotes the sample size.

Also, in the appendix to this chapter one will find a glossary of the terms that are followed in the text by an asterisk.

0.1 The Kernel Method

The kernel method was introduced and developed by Rosenblatt (1956) and Parzen (1962). It gives a general way of estimating derivatives of functions by smoothing the first differences of a crude estimator of these functions (usually a step function), using a weight function called a kernel.

Let $f(\cdot) = F'(\cdot)$. Suppose $F(\cdot)$ has been estimated at n points $x_1, \ldots, x_n$ (the data points) by $F_n(\cdot)$. Then $f(\cdot)$ is estimated by $f_n(\cdot)$:

$$f_n(x) = \sum_{j=1}^{n} \left\{ \frac{1}{h_j} K\left(\frac{x - x_j}{h_j} \right) \left[F_n(x_j + 0) - F_n(x_j - 0) \right] \right\}$$

where — $K(\cdot)$, the kernel, and $\{h_j\}$ are chosen appropriately

— $F_n(x_j + 0) - F_n(x_j - 0)$ is the value of the jump of $F_n(\cdot)$ at x_j .

Parzen and Rosenblatt were interested in estimating a probability density function (*). Then Watson and Leadbetter (1964) applied indirectly the same technique to hazard functions (*) . It could also be used for estimating sparsity function (*) or intensity functions (*) of counting processes; these applications have not been studied yet.

0.2 The Quantile Expansion Method

This method is due to G. P. Sillitto (1969, 1971). It has to do with estimating the quantile function (*), using shifted Legendre orthogonal polynomials. From the derivative of the quantile function, the sparsity function, one can easily get the density function.

Let $F(\cdot)$ be a strictly-increasing continuous distribution function (*), and $Q(\cdot)$ the associated quantile function.

Let $q(\cdot) = Q'(\cdot)$ and $f(\cdot) = F'(\cdot)$. Then

$$f(x) = \frac{1}{q(F(x))} \quad \text{and} \quad q(t) = \frac{1}{f(Q(t))}$$

The quantile function is estimated by

$$\hat{Q}_m(t) = \sum_{j=1}^{m} (2j - 1)\, \hat{\lambda}_j\, P^*_{j-1}(t)$$

where — P^*_{j-1} is the shifted (to [0,1]) Legendre polynomial of degree $(j - 1)$

— $\hat{\lambda}_j$ is a linear combination of the order statistics.

One then computes $\hat{q}_m(\cdot)$ and estimates $f(\cdot)$ by $\hat{f}_m(\cdot)$.

$$\hat{f}_m(\hat{Q}_m(t)) = \frac{1}{\hat{q}_m(t)}$$

0.3 The Spline Method

In the practice of density estimation, there seems to be different approaches placed in this category. Boneva, Kendall and Stefanov's method (1971) is really a variant of the kernel method where the chosen kernel, called a deltaspline, satisfies certain extremal properties arrived at through an original application of the theory of splines.

More properly (as to the classification), one can smooth the empirical quantile function or the empirical distribution function using spline functions and then differentiate these smooth versions, as in Wahba (1971).

A general description of this second approach runs as follows: given an ordered set of knots $\{x_i\}_{i=1}^{k}$, a corresponding set of estimated functional values $\{y_i\}_{i=1}^{k}$, find a real-valued function $F(\cdot)$ such that

$$F(x_i) = y_i , \qquad i = 1,\ldots,k$$

and other appropriate conditions are satisfied. These other conditions define the class of spline functions to be used, e.g., cubic splines, splines under tension, ...

We have considered splines under tension where the extra condition is that given $\sigma > 0$ (the tension factor), $F''(\cdot) - \sigma^2 F(\cdot)$ varies linearly on each of the intervals $[x_i, x_{i+1}]$, $i = 1,\ldots,k-1$.

This description was borrowed from Cline (1974). The imposed linearity condition allows estimation of the $\{y_i\}_{i=1}^{k}$ by least-squares method.

0.4 The Weighted Fourier Series Method

Let $f(\cdot)$ be nonnegative and defined on $[-\pi,\pi]$. Then $\varphi(v) = \int_{-\pi}^{\pi} e^{ivx} f(x)\,dx$ is determined by its value at $v = 0,1,2,\ldots$. Now if $f(\cdot)$ is square-integrable (*), we have the following inversion formula:

$$f(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-ivx} \varphi(v)$$

If $f(\cdot)$ is unknown, we can estimate $\varphi(v)$ by $\varphi_n(v) = \int_{-\pi}^{\pi} e^{ivx} dF_n(x)$, where $F_n(\cdot)$ is an estimator of the cumulative of $f(\cdot)$, and then invert a weighted version of $\varphi_n(\cdot)$ to get $f_n(\cdot)$

$$f_n(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-ivx} w(v)\, \varphi_n(v)$$

where $w(v)$ goes to 0 as $|v|$ gets large. Watson (1969) has determined that the optimal weights are

$$w(v) = \frac{|\varphi(v)|^2}{|\varphi(v)|^2 + \frac{1}{n}\,(1 - |\varphi(v)|^2)}$$

Some have used 0 - 1 weights (e.g., Kronmal and Tarter (1968)). Thaler (1974) has modeled the optimal weights from the sample.

Another approach that we favor is to use truncation along with estimated optimal weights:

$$f_m(x) = \frac{1}{2\pi} \sum_{v=-m}^{m} e^{-ivx} w_n(v)\, \varphi_n(v)$$

0.5 The Autoregressive Method

We now come to the main object of this dissertation. The autoregressive method got its name from time series analysis.

Let $F(\cdot)$ be a bounded nondecreasing function defined on $[-\pi,\pi]$. Let $R(\cdot)$ be the Fourier-Stieltjes transform of $F(\cdot)$,

$$R(v) = \int_{-\pi}^{\pi} e^{ivx}\, dF(x) , \qquad |v| = 0,1,2,\ldots$$

Solve the following system (Yule-Walker equations):

$$\begin{bmatrix} R(0) & R(1) & \cdots & R(m-1) \\ R(-1) & R(0) & \cdots & R(m-2) \\ & \vdots & & \\ R(-m+1) & R(-m+2) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} \alpha_{1m} \\ \alpha_{2m} \\ \vdots \\ \alpha_{mm} \end{bmatrix} = - \begin{bmatrix} R(-1) \\ R(-2) \\ \vdots \\ R(-m) \end{bmatrix} .$$

This can be seen either as fitting an m^{th} order autoregressive scheme (*) from the Yule-Walker equations involving a complex stationary covariance function $R(\cdot)$, or as building a set of orthogonal polynomials on the unit circle, with basis the complex exponentials and inner product defined by

$$(g,h) = \int_{-\pi}^{\pi} g(e^{ix})\, \overline{h(e^{ix})}\, dF(x)$$

The autoregressive approximator $f_m(\cdot)$ is given by

$$f_m(x) = \frac{K_m}{2\pi} \cdot \frac{1}{|1 + \alpha_{1m}e^{ix} + \alpha_{2m}e^{2ix} + \ldots + \alpha_{mm}e^{mix}|^2}$$

and is such that

$$\int_{-\pi}^{\pi} e^{ivx} f_m(x)\, dx = R(v) , \qquad |v| = 0,1,\ldots,m$$

Thus $f_m(\cdot)$ is approximating $\frac{dF(\cdot)}{dx}$ in a certain way.

The autoregressive estimator is obtained in the same way, except that $R(\cdot)$ has to be estimated first by $R_n(\cdot)$.

Appendix

0.A.1 Glossary of Terms

1 - A <u>distribution function</u> $F(\cdot)$ is a nondecreasing function continuous from the right and such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$. We will usually consider that $F(\cdot)$ is absolutely continuous with respect to Lebesgue measure, i.e., we will assume the existence of a function $f(\cdot)$, called a <u>density function</u>, such that $F(x) = \int_{-\infty}^{x} f(u)\, du$. Then, $f(\cdot) = F'(\cdot)$ a.e. Sometimes, we assume that $f(\cdot)$ is <u>square integrable</u>, i.e. $\int_{-\infty}^{\infty} |f(u)|^2 du < \infty$.

2 - Suppose that the distribution function $F(\cdot)$ is strictly increasing and absolutely continuous. Then we can define the functional inverse of $F(\cdot)$, denoted $Q(\cdot)$, called the <u>quantile function</u>. $Q(\cdot)$ is defined on $[0,1]$ and if $t = F(x)$, then $Q(t) = x$. Under the preceding conditions, there exists a function $q(\cdot)$, called the <u>sparsity function</u>, such that $q(\cdot) = Q'(\cdot)$. Now if we let $f(\cdot) = F'(\cdot)$, we will have the following relations:

$$f(x) = \frac{1}{q(F(x))} \qquad \text{and} \qquad q(t) = \frac{1}{f(Q(t))}$$

Tukey may have been the first one to use the term "sparsity."

3 - Now let $F(\cdot)$ be an absolutely continuous distribution function defined on $[0,\infty)$. The <u>hazard function</u> $h(\cdot)$ is defined as

$$h(x) = \frac{-d}{dx} \log(1 - F(x)) = \frac{f(x)}{1 - F(x)}$$

It is interpreted as an instantaneous failure rate

$$\lim_{dx \to 0} \frac{P(x < T \leq x + dx \mid T > x)}{dx} = h(x)$$

where T is a random variable with distribution $F(\cdot)$.

4 - For a counting process $N(t)$, we can define the concept of an intensity function $\lambda(\cdot)$ as

$$\lambda(t) = \frac{d}{dt} E[N(t)]$$

where $E[\cdot]$ denotes the expectation of a random variable.

5 - An m^{th} order autoregressive scheme is a stochastic process satisfying the following difference equation

$$X(t) + \alpha_{1m} X(t-1) + \ldots + \alpha_{mm} X(t-m) = \varepsilon(t)$$

where the $\varepsilon(\cdot)$'s are uncorrelated.

PART I - EMPIRICAL RESULTS

When one is presented with a new and radically different way of doing something for which there exists already quite a few techniques, one is naturally cautious. There is a natural principle of economy that needs to be respected before one may yield his approval. In the next two chapters we try to demonstrate in the most practical way that the autoregressive method deserves to become a standard technique of non-parametric curve estimation.

In the genesis of this work, the contents of these first two chapters opened the way to several questions to which much attention will be devoted in the second part. Also, in all fairness to the competing techniques, we include some practical suggestions that make them perform better.

CHAPTER 1

APPROXIMATING DENSITIES AND HAZARD FUNCTIONS

Most of the time, people use simulation to validate a new method. In the words of Boneva, Kendall and Stefanov (1971, p. 1): "We have for the present worked chiefly with simulated material for the excellent reason that what is essentially a diagnostic aid is most severely tested when one knows what the answer should be." In simulation, the answer is not exactly known. As a matter of fact, the sampling variability of the simulation process can yield "bad" samples. Should we feel free to sample until we get the "right" answer? This liberty is not available in the real data case. What can we do?

It depends on what the method can do. If the method can be used in a non-stochastic context, then it is possible to validate it without reference to sampling fluctuations. We call this process a first-hand validation. If the method is inseparable from the stochastic context, we may obtain consensus validation by comparing it to other methods on the same data. This is referred to as second-hand validation.

Consider the kernel method for instance. A first-hand validation would require n exact $F(\cdot)$ values. This would not be a very informative validation since we know that if n is large enough, even linear interpolation would give us a good fit for the derivative. The same can be said of the spline method.

In the autoregressive method, a first-hand validation requires knowledge of the true $R(\cdot)$ sequence. How many elements of the sequence do we need to get an approximator $f_m(\cdot)$ close enough to $f(\cdot)$? Now

this question remains meaningful in the real data case because the sample size n does not compel us to choose any order m. So any insight we gain from the validation process becomes handy when we are confronted with real data. The same can be said of weighted Fourier series.

Another reason to go through this process has to do with the structure of the estimation problem. Suppose we want to estimate a function $f(\cdot)$ and $f(\cdot)$ can be expressed in a form suitable for approximation by functions $\{f_m(\cdot)\}_{m=1}^{\infty}$. We might then estimate $f_m(\cdot)$ by $\hat{f}_m(\cdot)$. Now we have that

$$|f(\cdot) - \hat{f}_m(\cdot)|^2 = |[f(\cdot) - f_m(\cdot)] + [f_m(\cdot) - \hat{f}_m(\cdot)]|^2$$

The validation process can be of much help in the study and control of the bias function $b_m(\cdot)$:

$$b_m(\cdot) = f(\cdot) - f_m(\cdot)$$

In this chapter the emphasis is on validating the autoregressive method, though we will make some comparisons with the weighted Fourier series method.

In the probability density case, the function $R(\cdot)$ is simply the characteristic function. Accordingly we shall adopt the more usual notation $\varphi(\cdot)$ in the ensuing discussion.

1.1 The Idea of Truncating

A natural choice to start with is the Cauchy density

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{(1+x^2)} = \frac{1}{\pi} \cdot \frac{1}{|1+ix|^2}$$

with characteristic function $\varphi(v) = e^{-|v|}$, as our approximators are of the form

$$f_m(x) = \frac{K_m}{2\pi} \cdot \frac{1}{\left|1+\sum_{j=1}^{m} \alpha_{jm} e^{ijx}\right|^2}$$

Figure 1 reproduces the first picture we obtained. Obviously something had gone wrong. But then, our approximators are only defined on $[-\pi,\pi]$: more precisely they are periodic with period 2π ; and the interval $[-\pi,\pi]$ contains only about 80% of the Cauchy distribution. Moreover, even though the truncated density $f_T(\cdot)$ is proportional to the complete density $f(\cdot)$ on the interval of truncation T , it is zero outside that interval. That is

$$f_T(x) = \begin{cases} \dfrac{f(x)}{\int_T f(u)\,du} & , \quad x \in T \\ 0 & , \quad x \notin T \end{cases}$$

Thus the Fourier transform of $f_T(\cdot)$, $\varphi_T(\cdot)$, is not proportional to the Fourier transform of $f(\cdot)$, $\varphi(\cdot)$.

Let T be the interval $[-\pi,\pi]$. When $f(\cdot)$ is known, we can evaluate $\varphi_T(\cdot)$ directly using the Fast Fourier transform technique. If only $\varphi(\cdot)$ is known and is integrable, then

$$\varphi_T(v) = \frac{K}{\pi} \int_{-\infty}^{\infty} \varphi(\omega) \frac{\sin \pi(\omega - v)}{\omega - v} d\omega$$

where

$$K^{-1} = \int_{-\infty}^{\infty} \varphi(\omega) \frac{\sin \pi\omega}{\omega} d\omega$$

Using this correction, we produced Fig. 2 and 3 that show a remarkable fit. The distortion is not as important when most of the area is contained in the interval T.

What happens when the density is defined only on a subinterval of $[-\pi,\pi]$? Consider for example the uniform density on $[-1,1]$ with characteristic function $\varphi(v) = \frac{\sin v}{v}$. Figure 4 shows a wild behavior that is corrected only when the uniform density is made to fill the whole interval $[-\pi,\pi]$ as then $\varphi(0) = 1$ and $\varphi(v) = 0$, $v \neq 0$, which yields $f_m(\cdot) = \frac{1}{2\pi}$ for all values of m.

Note: The symbol o represents the true curve and the solid line the fitted one, unless otherwise noted.

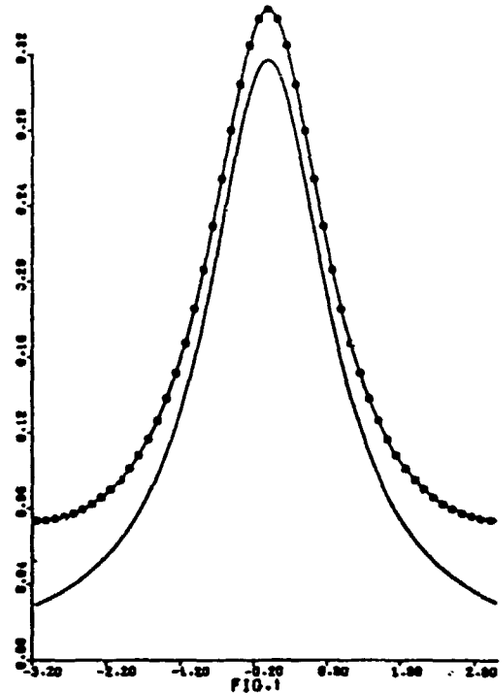


FIG.1

Cauchy not truncated
o : fitted curve

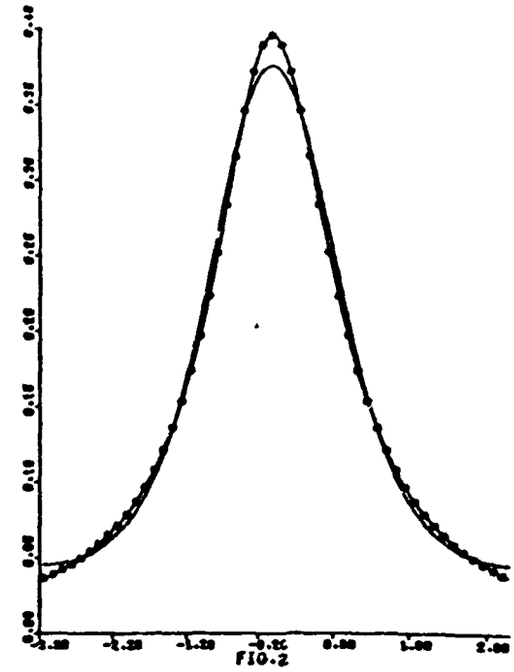


FIG.2

Cauchy order 2

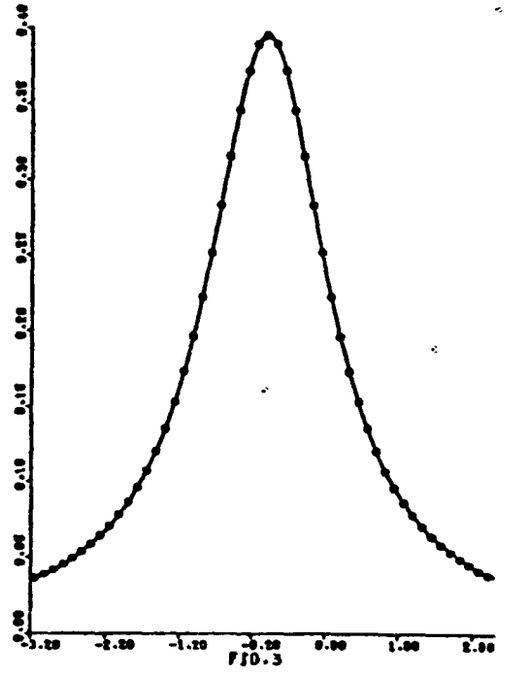


FIG.3

Cauchy order 8

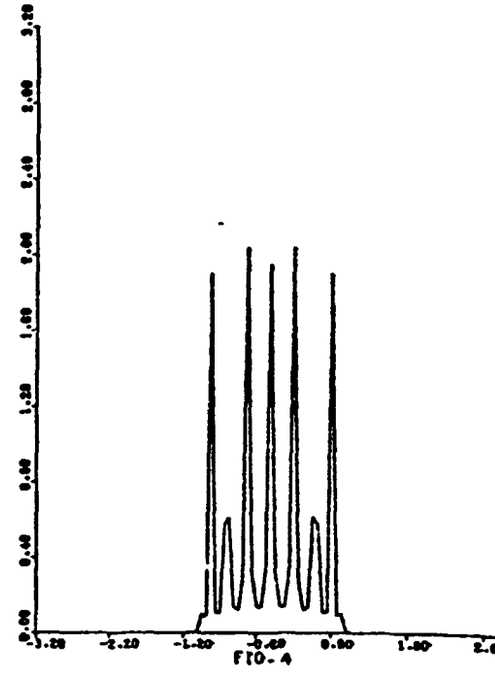


FIG.4

Uniform (-1,1) order 9

1.2 The Idea of Averaging

The wild behavior depicted in Figure 4 is more or less typical. This effect wears off quite rapidly in certain cases, but in others it persists. This difference in behavior is intriguing. Let us see the evidence.

Consider the standard Gaussian density. Why pay much attention to the low order approximators (Fig. 5 and 6) as convergence is still taking place (Fig. 7)? We could disregard this as an amusing oddity that unimodality and bimodality alternate. But when we shift the density by a small amount to a $N(0.5,1)$, the oddity becomes alarming because it persists and is aggravated (Fig. 8 and 9).

Now let us superimpose these pictures (Fig. 10 and 11). We get the remarkable fact that successive orders complement each other, so that if we average them (Fig. 12), the bias is greatly reduced.

This is not a problem of stability of the algorithm used, as the phenomenon is observed both with real and complex characteristic functions. More likely, this is related to the very structure of the approximator. We will have to examine this in Part II in our study of the bias function.

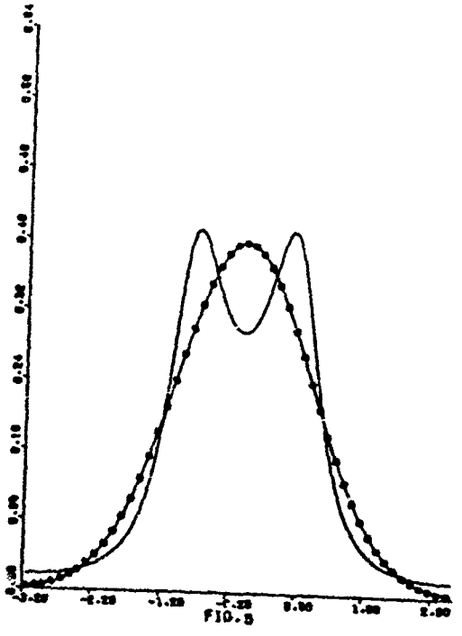


FIG. 5

Normal (0,1) order 2

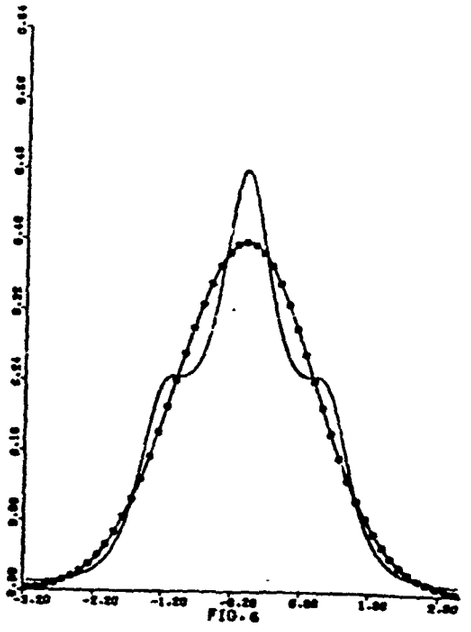


FIG. 6

Normal (0,1) order 3

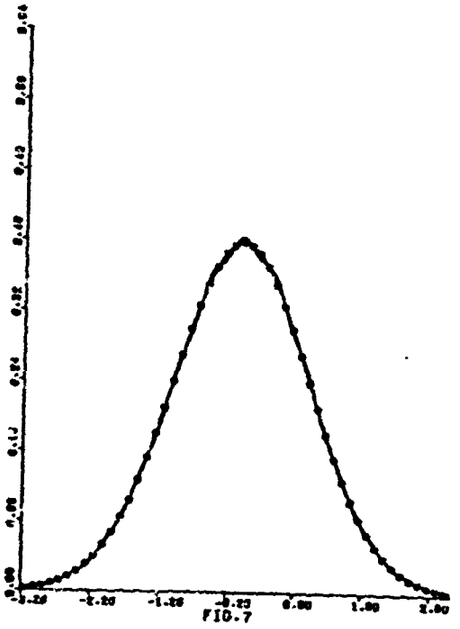


FIG. 7

Normal (0,1) order 11

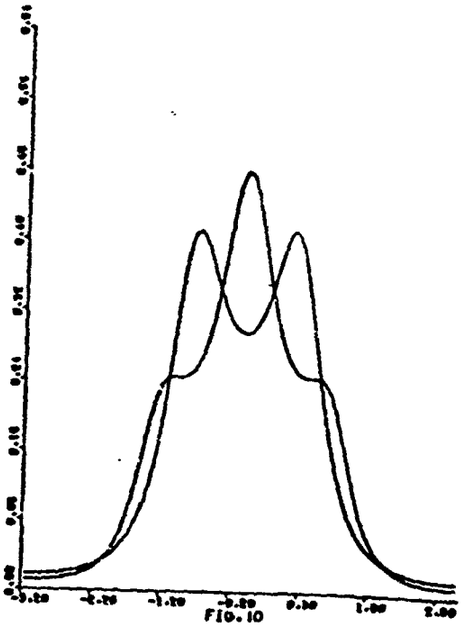


FIG. 10

Superposition of Fig. 5 and 6

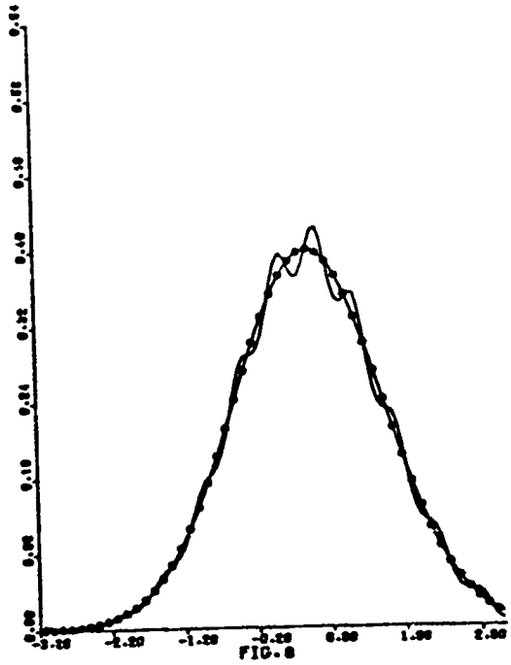


FIG. 8

Normal (0.5,1) order 10

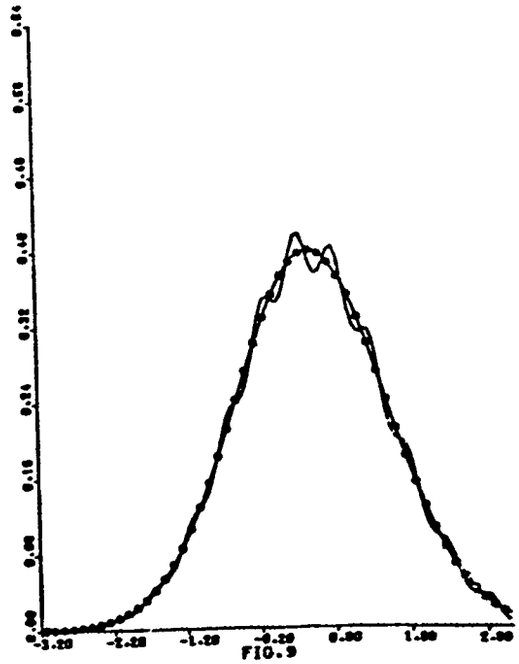


FIG. 9

Normal (0.5,1) order 11

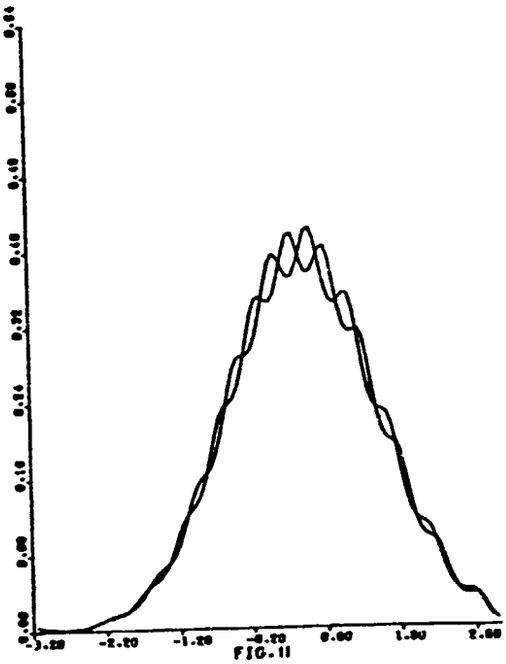


FIG. 11

Superposition of Fig. 8 and 9

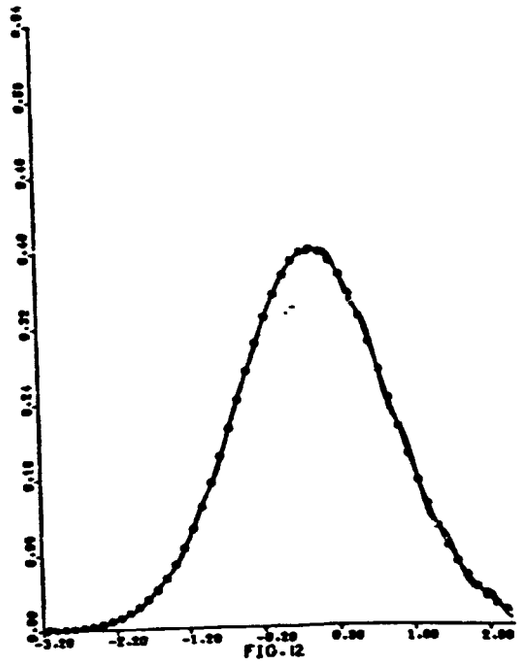


FIG. 12

Averaging of Fig. 8 and 9

1.3 The Idea of Symmetrizing

The next important class of functions to be used in our validation study is the class of gamma densities (exponential, chi-square, ...)

The exponential density, with its discontinuity at the origin, was expected to be difficult to approximate mainly because the approximators are periodic ($f_m(-\pi) = f_m(\pi)$) . Figures 13 and 14 illustrate our fears. One way to compensate the basic disequilibrium of the tails is to symmetrize the density so that both tails are equal.

We would like to point out that symmetry is not what matters but rather that the tails be comparable, as our approximation of non-central Gaussian densities show (Fig. 12).

By symmetrizing the exponential density, we get the Laplace (or double-exponential) density, and the discontinuity at the origin is now shifted down to the first derivative. In Figure 15 we see that the peak has been somewhat smoothed, but the fit is really good. The right half gives us a nice exponential curve.

The chi-square densities with larger degrees of freedom do not present any discontinuity, but their left and right-hand tails are very different. We shall compare symmetrization to scaling.

Let Y be a random variable distributed as a chi-square with four degrees of freedom. Let $X = 0.5Y$. The density of X is

$$f(x) = x e^{-x} \quad .$$

We symmetrize it to $f_x(x) = 0.5\,|x|\,e^{-|x|}$. Let us examine the results.

By comparing Figures 16 and 17, we see that symmetrization produces a smoother approximator than non-symmetrization for a given order. The same holds true after averaging two consecutive orders (Fig. 19 and 20). But when we symmetrize, we usually have to go to higher orders to get a better fit (Fig. 18). It is also possible to average more orders to get better results (compare Fig. 21 with Fig. 19).

Let us now look at the density of $X = 0.25Y$, i.e., $f(x) = 4x\,e^{-2x}$. The right extremity is at about the same height as the left. The approximators now perform well at both tails (Fig. 22), illustrating that symmetrization is not what we really need when the mode is more centrally located.

Note that symmetrization here is not taken in the same sense as in Feller (1966). The difference is apparent at the characteristic function level. Feller's procedure transforms the characteristic function into its square modulus. Our procedure transforms it into its real part.

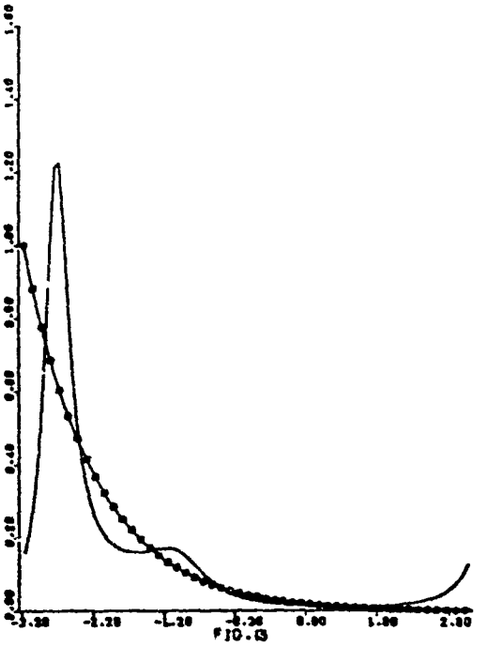


FIG.13

Exponential order 2

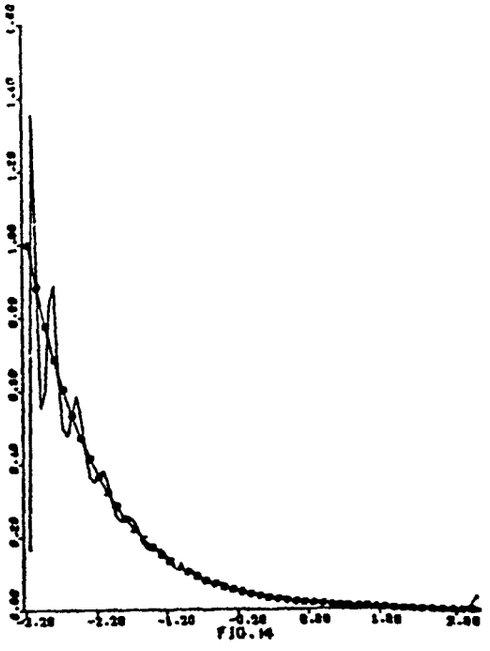


FIG.14

Exponential order 14

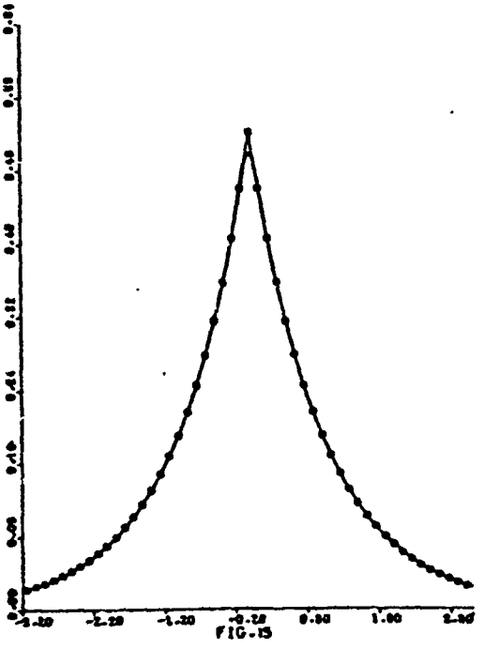


FIG.15

Laplace order 11

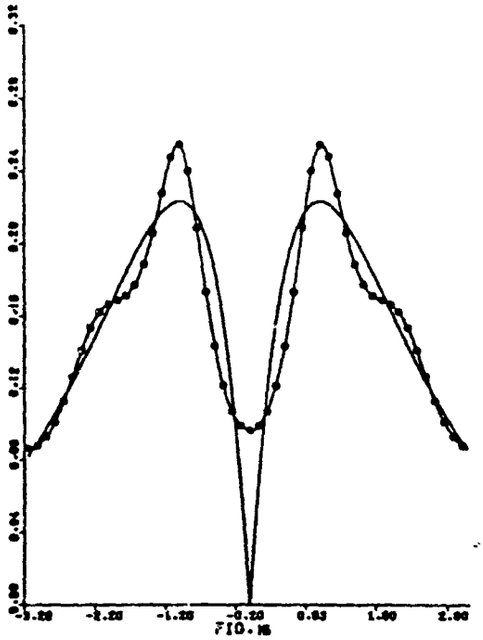


FIG. 16

Symmetric $0.5\ \chi^2$ (4 d.f.) order 4
o : fitted curve

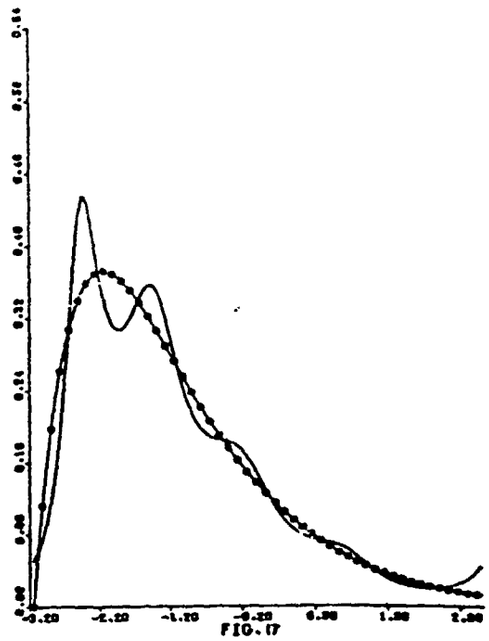


FIG. 17

$0.5\ \chi^2$ (4 d.f.) order 4

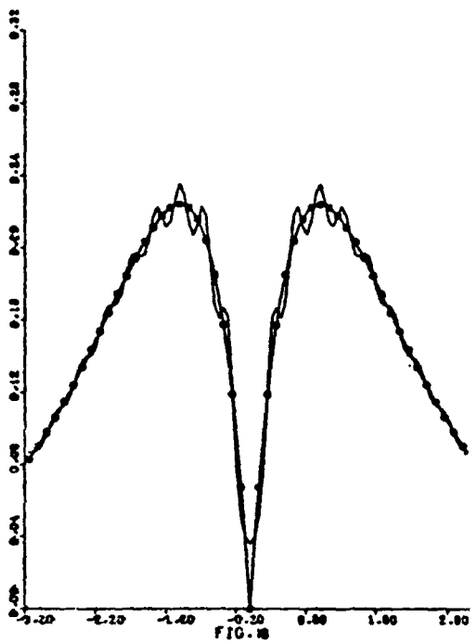


FIG. 18

Symmetric $0.5\ \chi^2$ (4 d.f.) order 17

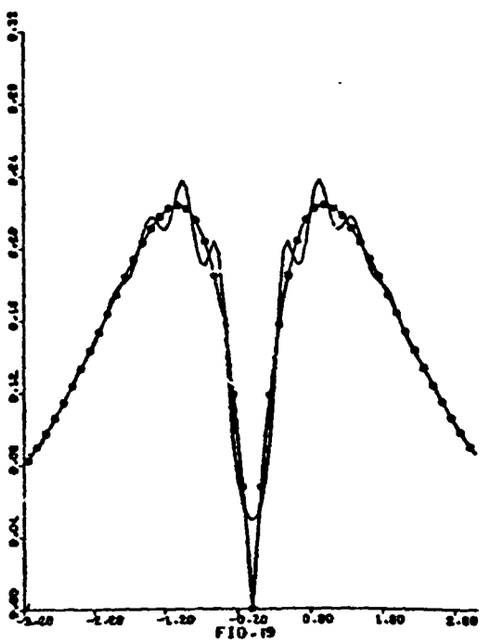
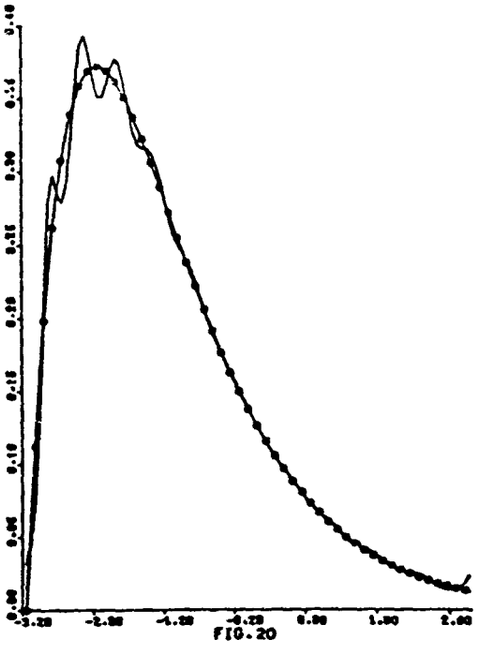
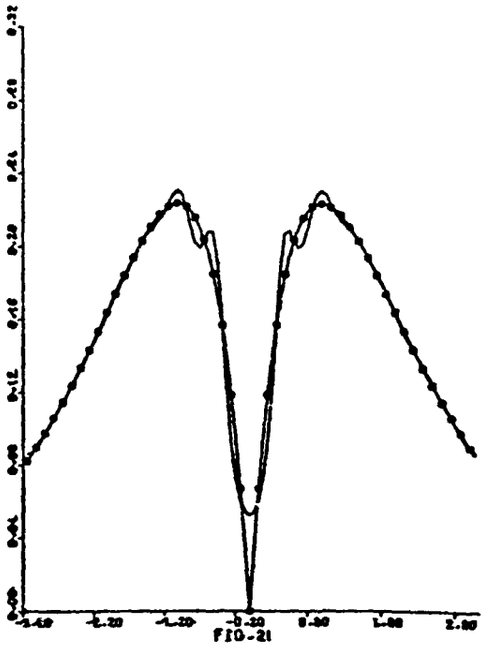


FIG. 19

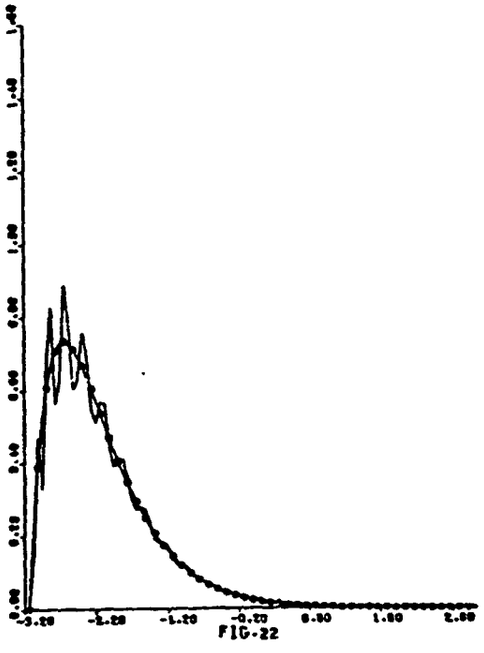
Symmetric $0.5\ \chi^2$ (4 d.f.) averaging orders 11 and 12



0.5 χ^2 (4 d.f.)
averaging orders 10 and 11



Symmetric 0.5 χ^2 (4 d.f.)
averaging orders 9, 10, 11 and 12



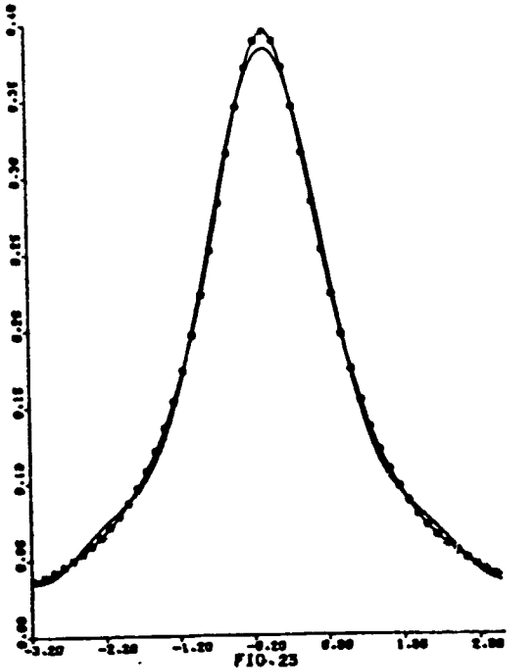
0.25 χ^2 (4 d.f.) order 19

1.4 Comparison with the Weighted Fourier Series Method

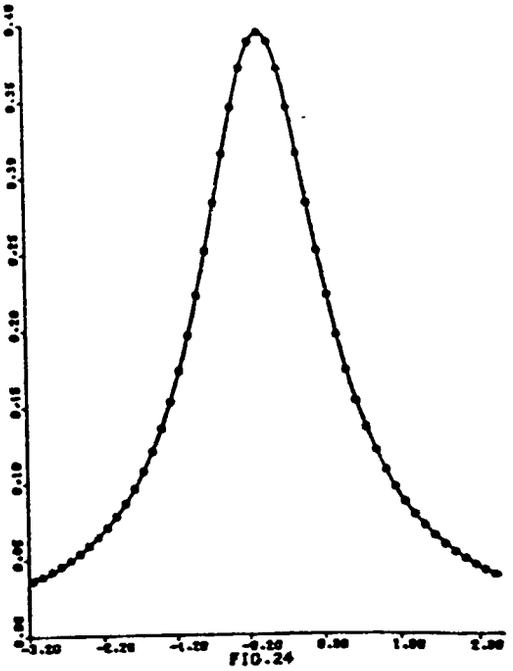
We first note that, in the case of validation, the weighted Fourier series method uses only 0 - 1 weights, i.e., the density is approximated by successive truncated Fourier series. This method requires as in the autoregressive method, that we truncate the density to $[-\pi,\pi]$. We also note tnat this is the standard way to invert a known Fourier transform. Thus, it should perform rather well.

For the Cauchy density, there is almost no difference between the two methods (Fig. 23 and 24 compared to Fig. 2 and 3). But in the case of the triangular density on $[-\pi,\pi]$, the autoregressive approximator wobbles about the true density (Fig. 25); averaging orders 8 and 9 reduces the bias pretty much (Fig. 26). But the truncated Fourier series is right on the target (Fig. 27).

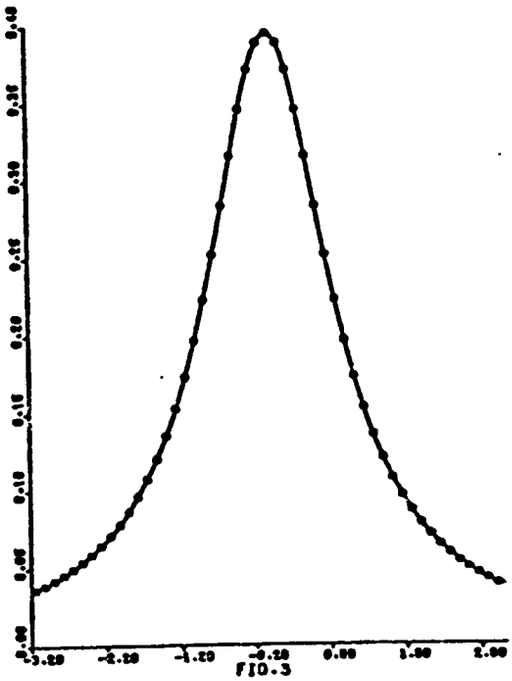
In the exponential case, we notice the same "odd-even" effect (Fig. 28 and 29). But there are differences: both tails are badly approximated (compared with Fig. 14) and the approximators become negative. Symmetrizing improves the matter (Fig. 30) as in the autoregressive method (Fig. 15).



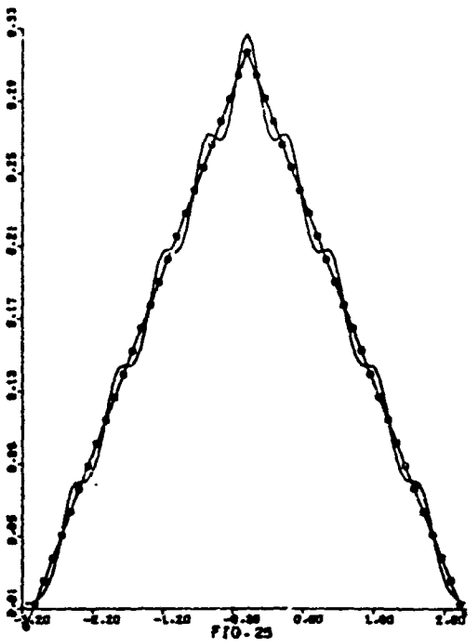
Cauchy (Fourier inversion) order 2



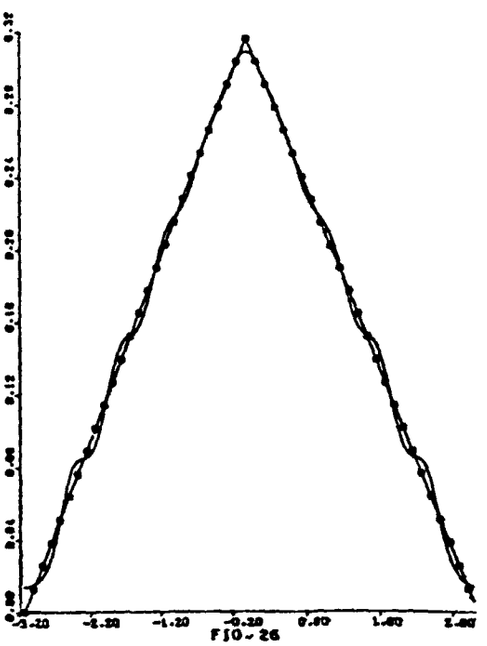
Cauchy (Fourier inversion) order 5



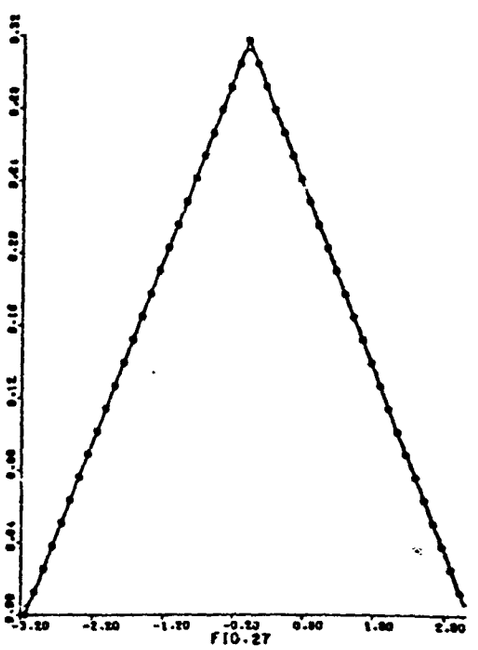
Cauchy (Autoregressive method) order 8



Triangular order 9



Triangular averaging orders 8 and 9



Triangular (Fourier inversion) order 12

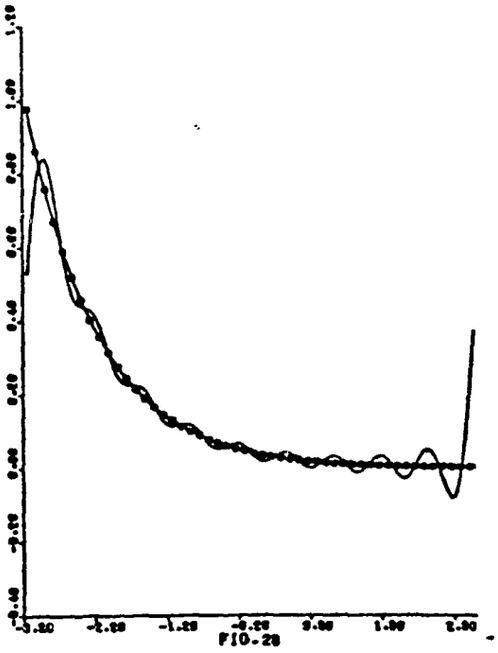


FIG.28

Exponential (Fourier inversion)
ier 9

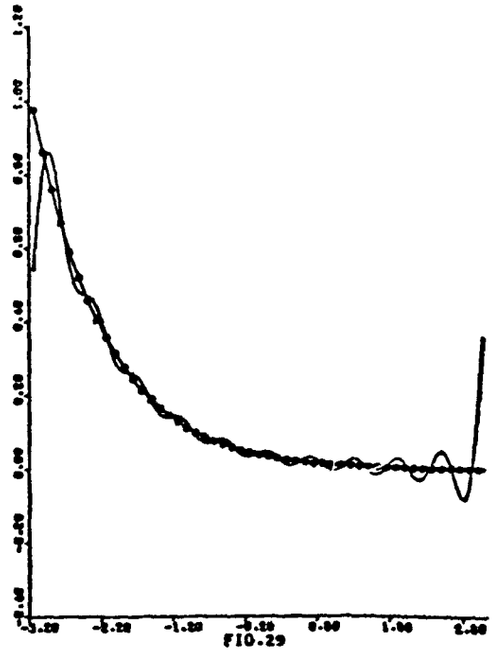


FIG.29

Exponential (Fourier inversion)
order 10

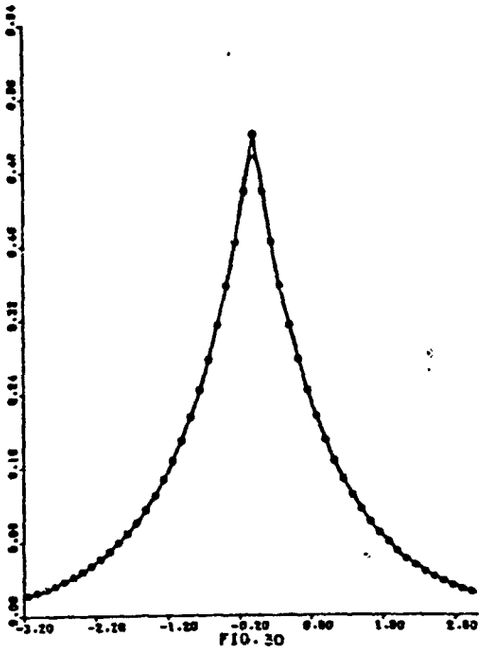


FIG.30

Laplace (Fourier inversion)
order 13

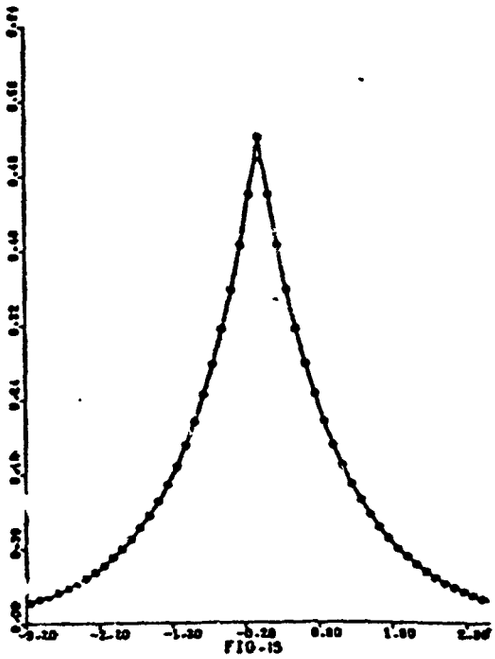


FIG-15

Laplace (Autoregressive method)
order 11

1.5 Approximation of a Hazard Function

When $R(\bullet)$ is taken as the Fourier transform of $\log(1 - F(\bullet))^{-1}$, the autoregressive method produces approximators of the hazard function related to the distribution $F(\bullet)$. In view of our previous work, we did not feel that it was necessary to include more than one example, which we have taken to be the hazard function of a truncated exponential (Fig. 31 and 32).

From the hazard function, it is possible to recover the original density as follows:

$$f(x) = h(x) \exp(-\int_0^x h(u)\, du)$$

When we apply this transformation to our hazard approximators, we obtain the density approximators represented in Figures 33 and 34. (Note we modified the hazard approximators at the origin to be 1.)

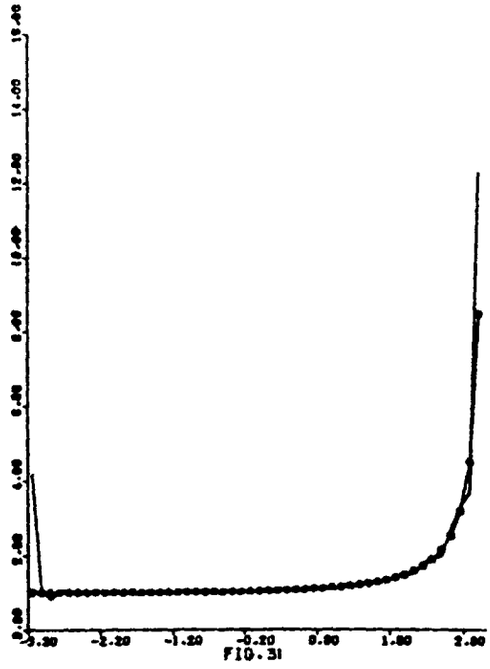


FIG. 31

Truncated exponential hazard order 16

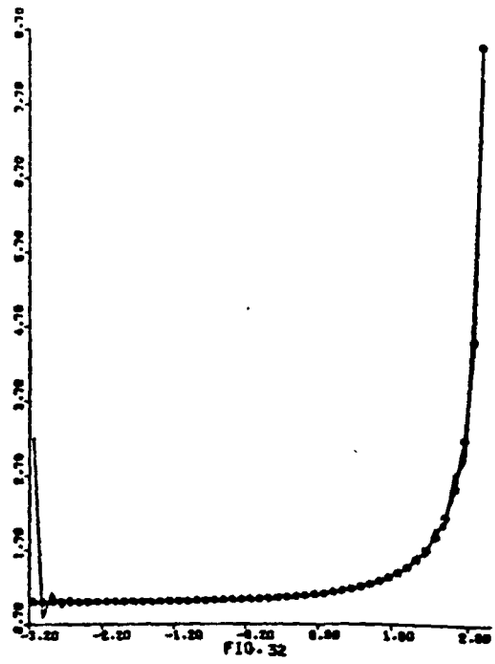


FIG. 32

Truncated exponential hazard order 26

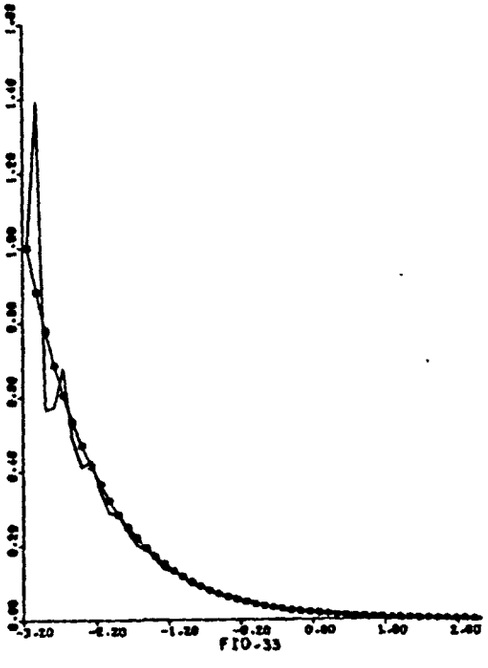


FIG. 33

Exponential from hazard order 16 modified at the origin

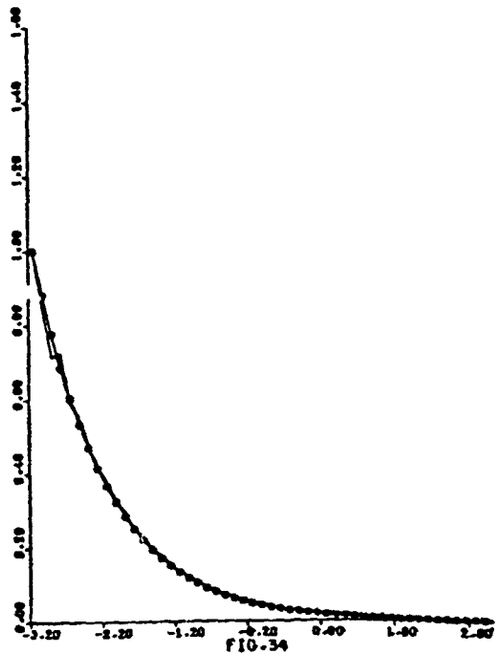


FIG. 34

Exponential from hazard order 26 modified at the origin

1.6 The Chi-square case revisited

The chi-square densities having shown to be more difficult to approximate, they furnish a valid testing ground for some alternate representations of the densities. We have already mentioned the hazard representation and the excellent results it produced for the exponential. In the chi-square case we obtain Figures 35 and 36. The results are not so good.

There is another representation that has not been used very often, namely the sparsity representation mentioned in the Appendix O.A.1 (item 2). The pictures we obtain (Fig. 37 and 38) compare favorably to direct approximators of higher order (e.g., Fig. 20).

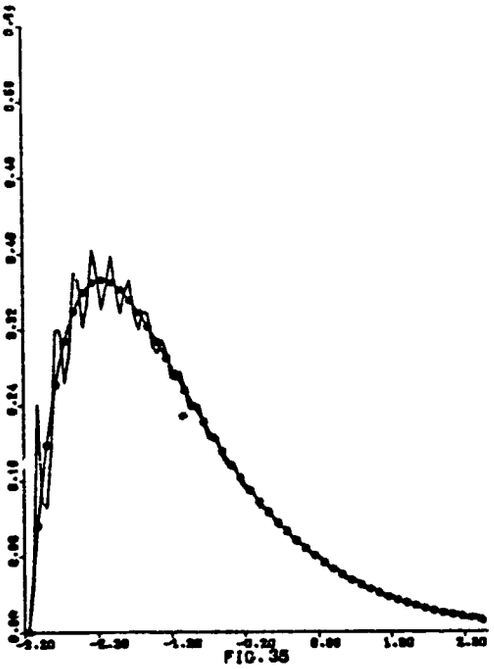


FIG. 35

χ^2 via hazard order 26

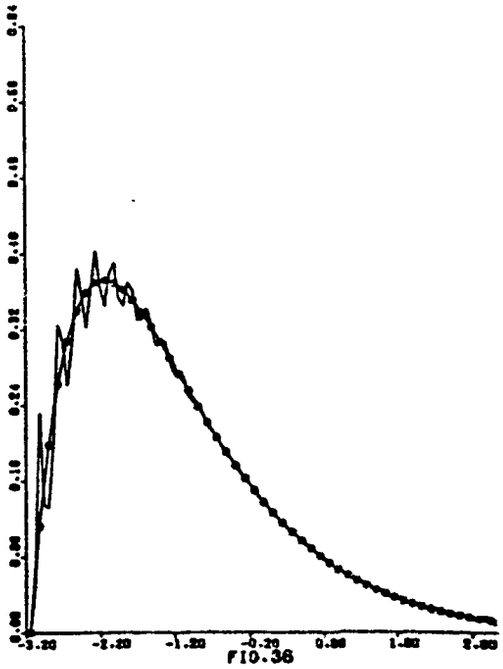


FIG. 36

χ^2 via hazard averaging orders 26 and 27

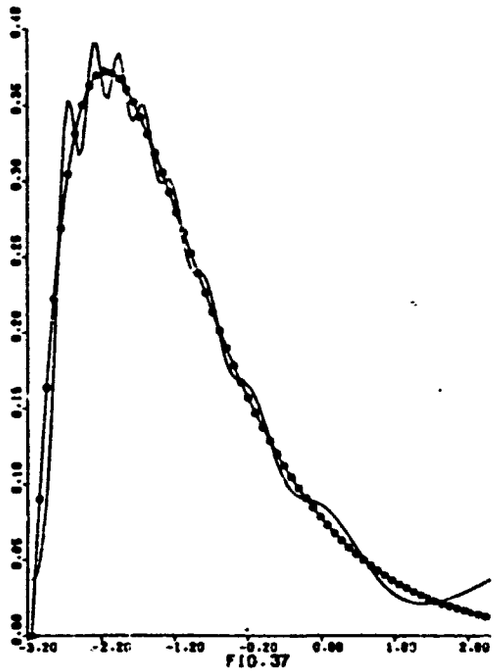


FIG. 37

χ^2 via sparsity order 8

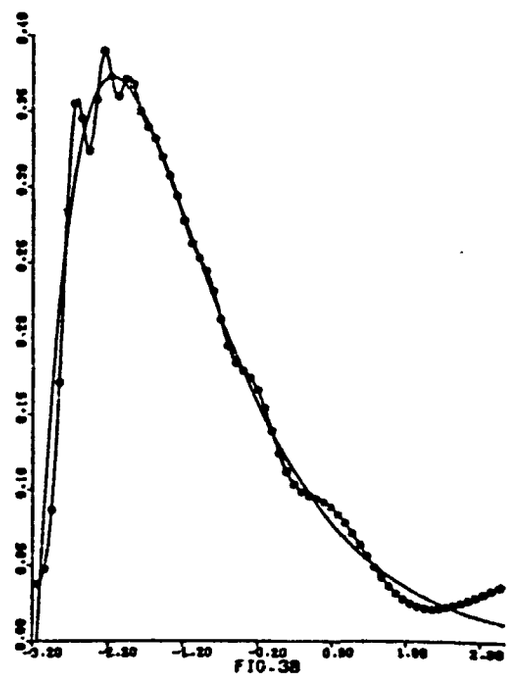


FIG. 38

χ^2 via sparsity averaging orders 7 and 8

1.7 A Look at the Output Parameters

We have left aside up to now all numerical considerations. To draw all the information contained in our validation work, there only remains to consider the output parameters that define our approximators: the coefficients $\{\alpha_{jm}\}_{j=1}^{m}$ and the proportionality factor K_m that normalizes the approximator to integrate to $R(0)$.

We consider only four cases to illustrate the points we want to make.

Table 1.1 exhibits the relation between the parameters and the kind of picture we get. K_m and the coefficients $\{\alpha_{jm}\}_{j=1}^{m}$ are converging nicely, the same way $f_m(\cdot)$ approximates the Cauchy density (Fig. 2 and 3).

Table 1.1 Some parameters of the Cauchy density autoregressive approximator

Order m	Coefficients α_{1m}	α_{2m}	Scale Factor K_m
1	-0.4838		0.7658
2	-0.5310	0.0974	0.7586
3	-0.5346	0.1175	0.7575
4	-0.5354	0.1198	0.7572
5	-0.5356	0.1203	0.7571
.			
.			
.			
11	-0.5358	0.1207	0.7570

When K_m goes to 0 , the coefficients do not converge, as in Table 1.2. There is also the danger that K_m will become negative, in which case the approximator $f_m(\cdot)$ is negative over the whole domain $[-\pi,\pi]$. Figure 4 is typical of the kind of picture associated with this behavior.

Table 1.2 Some parameters of the Uniform (-1,1) density autoregressive approximator

Order m	Coefficients α_{1m}	α_{2m}	Scale Factor K_m
1	-0.8414		0.2919
2	-1.5719	0.8681	0.0719
3	-2.3304	2.2415	0.0170
⋮			
11	-8.4770	34.5193	1.39×10^{-7}

Table 1.3 exemplifies slow convergence. In such a case we have found it helpful to average consecutive orders. The related pictures are shown in Figures 7, 8, 11 and 12.

Table 1.3 Some parameters of the Normal (0.5, 1) density autoregressive approximator

Order m	Coefficients α_{1m}	Coefficients α_{2m}	Scale Factor K_m
1	(-0.5414,0.2936)		0.6205
2	(-0.7592,0.4069)	(0.2230,-0.3302)	0.5220
3	(-0.8596,0.4538)	(0.3679,-0.5211)	0.4816
⋮			
9	(-1.0079,0.4961)	(0.6383,-0.7415)	0.4245
10	(-1.0152,0.4966)	(0.6535,-0.7471)	0.4218
11	(-1.0210,0.4969)	(0.6656,-0.7511)	0.4196

From Tables 1.4 and 1.5, there does not seem to be differences in convergence; this reaffirms our finding that there is no real gain in symmetrizing when the mode is relatively central. This is in sharp contrast with the exponential and Laplace cases in Tables 1.6 and 1.7.

Table 1.4 Some parameters of the 0.5 Chi-square (4) density autoregressive approximator

Order m	Coefficients α_{1m}	Coefficients α_{2m}	Scale Factor K_m
1	(-0.1860,-0.1224)		0.9503
2	(-0.2111,-0.1173)	(0.1069,0.0426)	0.9377
3	(-0.2207,-0.1155)	(0.1264,0.0369)	0.9310
⋮			
10	(-0.2461,-0.1142)	(0.1614,0.0347)	0.9097
11	(-0.2482,-0.1143)	(0.1642,0.0350)	0.9078

Table 1.5 Some parameters of the symmetrized Chi-square (4) density autoregressive approximator

Order m	Coefficients α_{1m}	Coefficients α_{2m}	Scale Factor K_m
1	-0.0950		0.9909
2	-0.1138	0.1970	0.9524
3	-0.1065	0.1928	0.9511
⋮			
18	-0.0854	0.2432	0.9162
⋮			
29	-0.0837	0.2449	0.9140

Table 1.6 Some parameters of the exponential density autoregressive approximator

Order m	Coefficients α_{1m}	Coefficients α_{2m}	Scale Factor K_m
1	(0.4998,-0.5001)		0.4999
2	(0.3997,-0.8001)	(-0.4001,-0.1997)	0.3999
3	(0.2996,-0.9001)	(-0.6000, 0.0004)	0.3599
⋮			
13	(0.0782,-0.9936)	(-0.5703, 0.4086)	0.2936
14	(0.0735,-0.9943)	(-0.5664, 0.4149)	0.2923

Table 1.7 Some parameters of the Laplace (symmetrized exponential) density autoregressive approximator

Order m	Coefficients α_{1m}	Coefficients α_{2m}	Scale Factor K_m
1	-0.5456		0.7023
2	-0.6209	0.1381	0.6889
3	-0.6331	0.1928	0.6835
⋮			
10	-0.6370	0.2027	0.6820
11	-0.6370	0.2028	0.6820

1.8 Conclusion

1. It is important to realize that the autoregressive method approximates only functions truncated to $[-\pi,\pi]$ and that the domain of definition of these functions should fill the whole interval.

2. There seems to be a structural "odd-even" effect that can be averaged out.

3. Symmetrizing improves the behavior of the approximators when the functions have a maximum at the left end of the domain.

4. In view of the periodicity of the approximators, they perform better when both ends of the functions are comparable.

5. Satisfying approximators are related to the convergence of the parameters.

APPENDIX

I.A.1 Sample Programs for Approximation

We include in this appendix three sample programs, one for each of the approaches we used.

Each program is divided in 3 parts:

I - Computing the R(•) sequence

II - Solving the Yule-Walker equations in AUTOREG. This subroutine can be found in the appendix to Chapter 3.

III - Computing the density.

In the direct approach, II and III are confounded.

1st program: Approximating a symmetrized chi-square.

At the beginning, A(•) contains the function to be approximated. Then, using two IMSL subroutines FFT2 and FFRDR2, we compute the R(•) sequence that is stored in A(•). This is the end of Part I.

In Part II, we solve the Yule-Walker equations and compute the approximator stored in F(•). NORM(•) contains the true truncated density being approximated (NORM(•) is used only for plotting purposes). We also average four consecutive orders, the average being stored in G(•).

2^{nd} program: Approximating the hazard of a chi-square.

The first two parts are as in the first program. At the end of Part II, F(•) contains the approximated hazard.

In Part III we reconstitute the density from the hazard function. At the end of Part III, F(•) contains the approximated density.

3^{rd} program: Approximating the sparsity of a chi-square.

In Part I, G(•) contains the chi-square density and CF(•) is the quantile function obtained by the trapezoidal rule. Then, we rescale CF(•) to be between $-\pi$ and π, using subroutines CENTER and XSCALE, and compute in FOURST I the R(•) sequence stored in A(•) :

$$R(v) = \int_{-\pi}^{\pi} e^{iv(2\pi \cdot F(x)-\pi)}\, dx = \int_{-\pi}^{\pi} e^{ivt}\, q\left(\frac{t+\pi}{2\pi}\right) dt$$

by letting $t = 2\pi \cdot F(x) - \pi$.

In Part II, we solve the Yule-Walker equations. F(•) now contains the sparsity.

Finally in Part III, we recover the density. Subroutine FOURSTI simply evaluates

$$\sum_{j=1}^{n} e^{iv\,X(j)} \cdot F(j) \quad , \qquad v = \frac{1}{L}, \frac{2}{L}, \ldots, \frac{M \cdot L}{L}$$

using the function CSREC , which is a recursion for cosines and sines:

```
FUNCTION CSREC(C1,C2,C3)
CS REC  =  C1 * C2 - C3
C3  =  C2
C2  =  CSREC
RETURN
END
```

```
      PROGRAM SWEEP(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      COMPLEX AA
      COMPLEX ALPHA(30),PHI(30),JH,KH
      INTEGER IWK(12)
      COMPLEX A(2048)
      DIMENSION G(102)
      DIMENSION X(102),F(102)
      REAL NORM(102)
      PI=4.*ATAN(1.0)
      TWOPI=2.*PI

      NI=2048
      FNI=NI
      DO 21 J=1,NI
      FREQ=TWOPI*(J-1)/FNI
      A(J)=CMPLX(ABS(FREQ-PI)*EXP(-ABS(FREQ-PI)),0.)
   21 CONTINUE
      CALL FFT2(A,11,IWK)
      CALL FFROR2(A,11,IWK)
      AA=A(1)
      M=15
      DO 3 J=1,M
      A(J)=A(J)*(-1)**(J-1)/A(1)
    3 CONTINUE
      AA=AA*TWOPI/FNI
      PRINT*,AA
      NP=100
      FNP=NP
      DO 13 I=1,NP
      X(I)=-PI+(I-1)*TWOPI/FNP
      G(I)=0.
      NORM(I)=ABS(X(I))*EXP(-ABS(X(I)))/AA
   13 CONTINUE


      DO 2 K=2,M
      CALL AUTOREG(A,K,M,NP,ALPHA,PHI,JH,KH,X,F)
      IF(K.NE.10.AND.K.NE.11.AND.K.NE.12.AND.K.NE.13) GO TO 2
      DO 24 I=1,NP
      G(I)=G(I)+0.25*F(I)
   24 CONTINUE


    2 CONTINUE

      STOP
      END
```

```
      PROGRAM SWEEP(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      COMPLEX A(2048)
      COMPLEX ALPHA(30),PHI(30),JH,KH
      INTEGER IWK(12)
      REAL NORM(250),X(250),F(250)
      DIMENSION G(330),CF(100)
      DIMENSION R(30),C(30)
      PI=4.*ATAN(1.0)
      TWOPI=2.*PI
      NI=2048
      FNI=NI
      TT=(TWOPI+1.)*EXP(-TWOPI)
      AA=1.-TT
      DO 21 J=1,NI
      FREQ=TWOPI*(J-1)/FNI
      A(J)=CMPLX(FREQ*EXP(-FREQ)/((FREQ+1.)*EXP(-FREQ)-TT),0.)
   21 CONTINUE
      CALL FFT2(A,11,IWK)
      CALL FFRDF2(A,11,IWK)
      M=27
      DO 3 J=1,M
      A(J)=A(J)*(-1)**(J-1)*TWOPI/FNI
    3 CONTINUE
      NP=150
      FNP=NP
      DO 13 I=1,NP
      X(I)=-PI+(I-1)*TWOPI/FNP
      NORM(I)=(X(I)+PI)*EXP(-(X(I)+PI))/AA
   13 CONTINUE
      H=PI/FNP

      DO 2 K=2,M
      CALL AUTOREG(A,K,M,NP,ALPHA,PHI,JH,KH,X,F)
      IF(K.NE.17.AND.K.NE.27) GO TO 2
      PRINT*,F

      F(1)=0.
      F(2)=0.05
      G(1)=0.
      DO 24 I=2,NP
      G(I)=G(I-1)+H*(F(I-1)+F(I))
   24 CONTINUE
      DO 25 I=1,NP
      F(I)=F(I)*EXP(-G(I))
   25 CONTINUE


      PRINT*,F
    2 CONTINUE

      STOP
      END
```

I

II

III

```
      PROGRAM SWEEP(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      COMPLEX  A(30)
      COMPLEX ALPHA(30),PHI(30), JH, KH

      REAL NORM(150),X(150),F(150)
      DIMENSION G(130),CF(130)
      DIMENSION R(30),C(30)

      PI=4.*ATAN(1.0)
      TWOPI=2.*PI
      NI=128
      FNI=NI
      H=PI/FNI
      DO 21 J=1,NI
      FREQ=TWOPI*(J-1)/FNI
      G(J)=FREQ*EXP(-FREQ)
   21 CONTINUE
      CF(1)=0.
      DO 4 I=2,NI
      CF(I)=CF(I-1)+(G(I-1)+G(I))*H
    4 CONTINUE
      AA=CF(NI)
      DO 5 I=1,NI
      CF(I)=CF(I)/CF(NI)
      G(I)=2.*H
    5 CONTINUE
      PRINT*,CF
      XMID=0.5
      XRNG=1.
      SCL=TWOPI/XRNG
      CALL CENTER(CF,CF,NI,XMID)
      CALL XSCALE(CF,CF,NI,SCL)
      PRINT*,CF
      M=10
      L=1
      CALL FOURSTI(CF,NI,R,C,M,L,G)
      PRINT*,CF
      DO 3 J=2,M
      L=J-1
      A(J)=CMPLX(R(L),C(L))
    3 CONTINUE
      A(1)=CMPLX(TWOPI,0.)
      NP=128
      FNP=NP
      PRINT*,A
      DO 13 I=1,NP
      X(I)=-PI+(I-1)*TWOPI/FNP
      NORM(I)=(X(I)+PI)*EXP(-(X(I)+PI))/AA
      G(I)=0.
   13 CONTINUE
      PRINT*,NORM
```

```
      DO 2 K=2,M
      CALL AUTGREG(A,K,M,NF,ALPHA,PHI,JH,KH,CF,F)
      IF(K.LT.8) GO TO 2
      WRITE(6,6) (ALPHA(I),I=1,K),KH
    6 FORMAT(//,1X,4(2F8.4,3X))
      IF(K.NE.8.AND.K.NE.9) GO TO 2
      DO 25 I=1,NF
      F(I)=1./(F(I)*TWOPI)
   25 CONTINUE
      PRINT*,F
      DO 24 I=1,NF
      G(I)=G(I)+0.5*F(I)
   24 CONTINUE


    2 CONTINUE

      STOP
      END


      SUBROUTINE FCURST1(X,N,CPHI,SPHI,M,L,F)
      DIMENSION X(N),F(N),CPHI(1),SPHI(1)
      PI=4.*ATAN(1.0)
      EL=FLOAT(L)
      ML=M*L
      DO 50 J=1,ML
      CPHI(J)=SPHI(J)=0.
   50 CONTINUE
      DO 100 I=1,N
      IF(ABS(X(I)).GT.PI) GO TO 104
      XX=X(I)/EL
      C1=COS(XX)
      C0=2.*C1
      C2=1.
      S1=SIN(XX)
      S2=0.
      CPHI(1)=CPHI(1)+C1*F(I)
      SPHI(1)=SPHI(1)+S1*F(I)
      DO 150 J=2,ML
      CPHI(J)=CPHI(J)+CSREC(C0,C1,C2)*F(I)
      SPHI(J)=SPHI(J)+CSREC(C0,S1,S2)*F(I)
  150 CONTINUE
  100 CONTINUE
      GO TO 5
  104 WRITE(6,103)
  103 FORMAT(1H1,2X,24HDATA NOT SCALED PROPERLY)
    5 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CENTER(X,Y,N,XMID)
      DIMENSION Y(1)
      DIMENSION X(1)
      XPT=XMID
      IF(N.GT.0) XPT=-XMID
      NN=IABS(N)
      DO 9 I=1,NN
    9 Y(I)=X(I)+XPT
      RETURN
      END
```

```
      SUBROUTINE XSCALE(X,Y,N,SCL)
      DIMENSION Y(1)
      DIMENSION X(1)
      PISCL=SCL
      IF(N.LT.0) PISCL=1./PISCL
      NN=IABS(N)
      DO 10 I=1,NN
   10 Y(I)=X(I)*PISCL
      RETURN
      END
```

CHAPTER 2

ESTIMATING DENSITIES AND HAZARD FUNCTIONS

The kind of validation that we have completed in chapter 1 is not sufficient for a method that is to be used in a statistical context. We have to confront it with real data. This testing can best be done by what was called before a second-hand validation.

Even though our main interest remains with the autoregressive method, we will have to take a longer look at the competing techniques that we mentioned in our introduction, examining critically different more or less inaccurate maps to recognize the ground we are standing on. We do not define with any more precision the tasks that these methods could be asked to perform. It will suffice for the moment to see how they describe the data. This is admittedly an incomplete assessment from which we shall not try to state definitive answers.

This critical examination will be done using a diversity of real data situations: approximately normal data, approximately exponential data and frequency data. But first we indicate how we have used the different methods.

2.1 Choice of Input Parameters

2.1.1 The Kernel Method

The kernel method contains two input parameters, namely the kernel function and the set $\{h_j\}_{j=1}^{n}$ of bandwidths.

We have used different kernels: Parzen kernel (Fig. 1, 2, 18, 30), Gaussian kernel (Fig. 3), naive kernel (Fig. 4). It is clear that the choice of a kernel is an important one. The Parzen or Gaussian kernels will produce in general smoother estimators.

We have always used constant bandwidths $h_j \equiv h$, even though there are algorithms of the nearest neighbor type to adapt the h_j's to each data point, following Loftsgaarden and Quesenberry (1965).

For the particular kernels we used, the optimal value of h is given by

$$h_{opt} = \left\{ \frac{f(x) \int K^2(y)\, dy}{4n[f''(x) \cdot \int y^2 K(y)\, dy]^2} \right\}^{1/5}$$

$$= \frac{[f(x)]^{1/5}}{[2f''(x)]^{2/5}} \cdot \left[\frac{KERFAC}{n}\right]^{1/5}$$

(Parzen (1962))

We have approximated this by

$$K \cdot STDEV \cdot \left(\frac{KERFAC}{n}\right)^{1/5}$$

where STDEV is the standard deviation of the sample. The proportionality constant K , if it is too large, yields very flat estimators, and if too small, very spiky. A value around 2/3 has worked very well. Note also that STDEV is very sensitive to outliers so that adjustments may have to be made.

2.1.2 The Quantile Expansion Method.

The only parameter here is the order of approximation. So we use the method iteratively, until we are satisfied.

2.1.3 The Spline Method.

We have usually fitted splines under tension to the empirical quantile function. The input parameters include the tension factor, the set of knots $\{x_i\}_{i=1}^{k}$ and the end-point conditions.

The choice of the knot points is very delicate and crucial. We usually started with 13 equidistant knots between 0 and 1 . Then we moved them around and reduced their number according to the pictures we were getting. This procedure is necessary when the estimated quantile function starts to decrease, contrary to the theory.

When the tension factor is less than .001 , the spline under tension is very much like the cubic spline; for values larger than 50 , it is a polygonal line.

The end-point conditions consist in this case of estimates of the first derivative of the quantile function at both ends, usually first differences.

2.1.4 The Weighted Fourier Series Method.

The input parameters here are the empirical Fourier transform, the weights and the order of approximation.

So again we use the method iteratively and estimate the optimal weights by using the empirical Fourier transform $\varphi_n(\cdot)$. Thaler (1974) has shown that for large v , $\varphi_n(v)$ is not a good estimate of $\varphi(v)$ in the sense that

$$\lim_{|v|\to\infty} \text{Var } \varphi_n(v) = \frac{1}{n}$$

But this should not cause us to worry as we usually need only small values of v , and we can check graphically (Fig. 9) that we are in a safe part of the domain by finding the value v_0 at which $|\varphi_n(v)|^2$ starts oscillating around $1/n$.

This procedure is simpler than Thaler's own proposal and it also performs better.

2.1.5 The Autoregressive Method.

The autoregressive method has the same kind of parameters as the weighted Fourier series method, minus the weights.

The Fourier transform $R(\cdot)$ is estimated by $R_n(\cdot)$ from the sample. Not only is the method used iteratively, but there is a recursive algorithm to go from one order to the next (see appendix 3.A.1).

2.2 Buffalo Snowfall Data

This set of data has been much studied in our department, more from the time series point of view. It consists of the 63 yearly values of snow precipitation, recorded to the nearest tenth of an inch, from 1910 to 1972. It was chosen to illustrate the response of the different methods to approximately normal data (see Fig. A1 and A2).

Table A

Buffalo Snowfall Data

Year	0	1	2	3	4	5	6	7	8	9
1910	126.4	82.4	78.1	51.1	90.9	76.2	104.5	87.4	110.5	25.0
1920	69.3	53.5	39.8	63.6	46.7	72.9	79.6	83.6	80.7	60.3
1930	79.0	74.4	49.6	54.7	71.8	49.1	103.9	51.6	82.4	83.6
1940	77.8	79.3	89.6	85.5	58.0	120.7	110.5	65.4	39.9	40.1
1950	88.7	71.4	83.0	55.9	89.9	84.8	105.2	113.7	124.7	114.5
1960	115.6	102.4	101.4	89.8	71.5	70.9	98.3	55.5	66.1	78.4
1970	120.5	97.0	110.0							

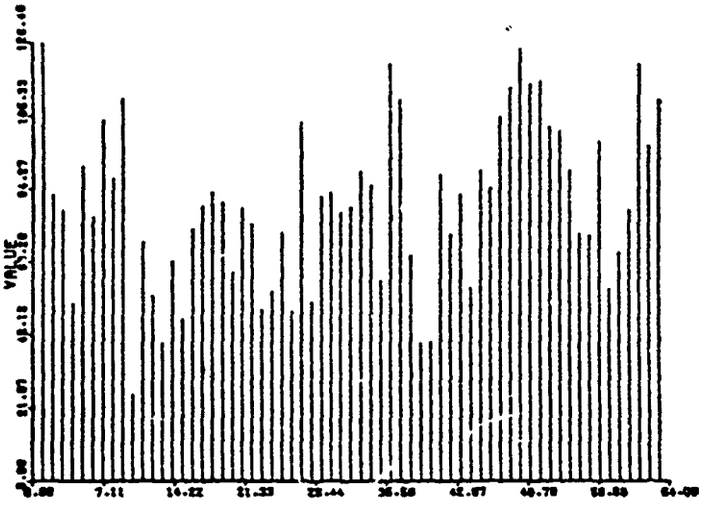


Figure A1
Buffalo Snowfall data

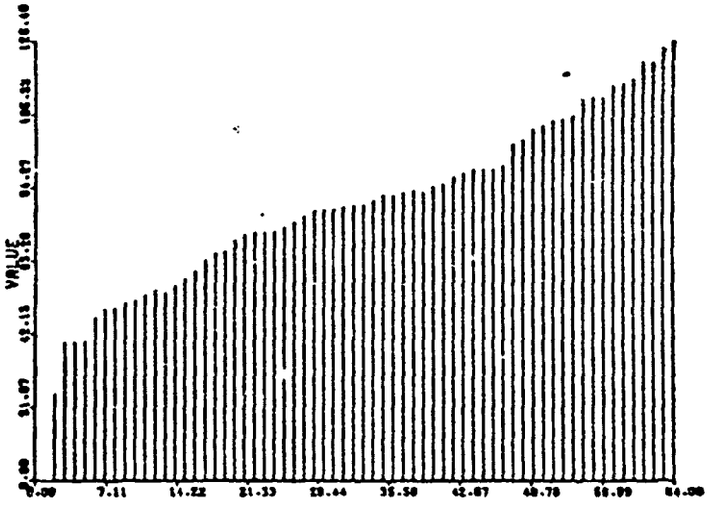


Figure A2
Same data ordered

2.2.1 The Kernel Method

We experiment first with different values of h. Figure 1 shows the effect of choosing h too large ($h = 27.75$) in comparison with Figure 2 ($h = 18.5$), where the kernel used is the Parzen kernel. Now we can compare the Gaussian kernel with the Parzen kernel: with the best choice of h, they yield the same estimate (Fig. 2 and 3). Thus the Parzen kernel is equivalent to the Gaussian kernel. The naive kernel yields spiky results even with the best choice of h (Figure 4). However the same basic shape can be distinguished. Note that this method does not impose any truncation on the data so that the tails always go to zero.

2.2.2 The Quantile Expansion Method

In Figure 5, the peaks are much sharper and more separated than in the previous pictures. It is at that order (order 8) that the three modes appeared for the first time.

2.2.3 The Spline Method

As in the previous case, the quantile function is smoothly estimated and then differentiated to produce the density estimator.

The effect of increasing the tension factor tenfold is pictured in Figure 6 ($\sigma = 1.5$) and Figure 7 ($\sigma = 15.$) . In Figure 8 we show the estimated quantile function that produced the estimate in Figure 6. It was based on 8 knot points located at 0., 0.06875, 0.25, 0.4, 0.625, 0.833, 0.916, 1.0 .

2.2.4 The Weighted Fourier Series Method

We determine first the value v_0 which is a kind of upper bound above which it would be "unsafe" to use the empirical characteristic function to estimate $\varphi(v)$. Figure 9 is a plot of $\log_{10} |\varphi_n(v)|^2$ vs. $\log_{10} v$ on a log-log scale. v_0 is such that $|\varphi_n(v)|^2$ oscillates around $1/n$ for $v > v_0$. Here $n = 63$. Thus we draw a horizontal line at $1/63 \cong 0.016$ to get that $v_0 > 9$.

We notice in Figure 10 that the location of the modes has not changed. The values of the two extreme modes are larger than in Figure 2 because of the truncation, the data having been rescaled to fill the interval $[-\pi,\pi]$ completely. The truncation is responsible too for the behavior at the left tail, which was also heavy in Figure 2.

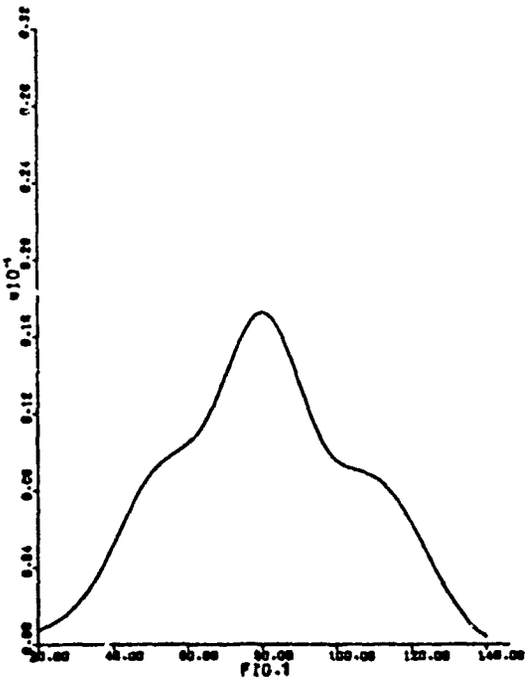


FIG.1

Parzen kernel, h = 27.75

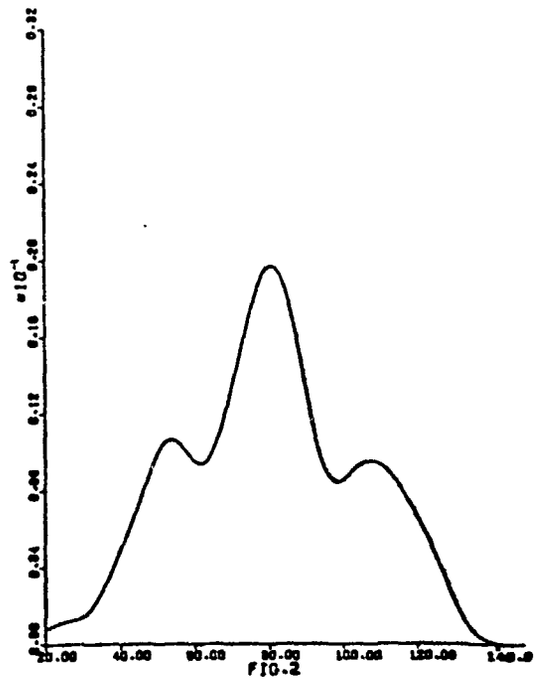


FIG.2

Parzen kernel, h = 18.5

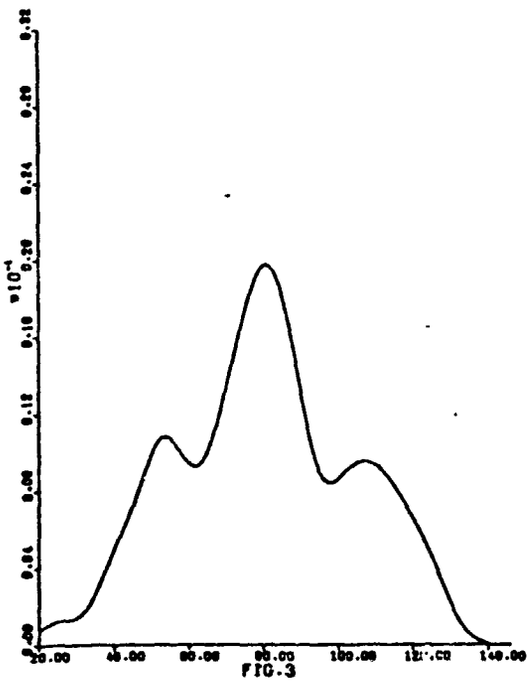


FIG.3

Gaussian kernel, h = 5.36

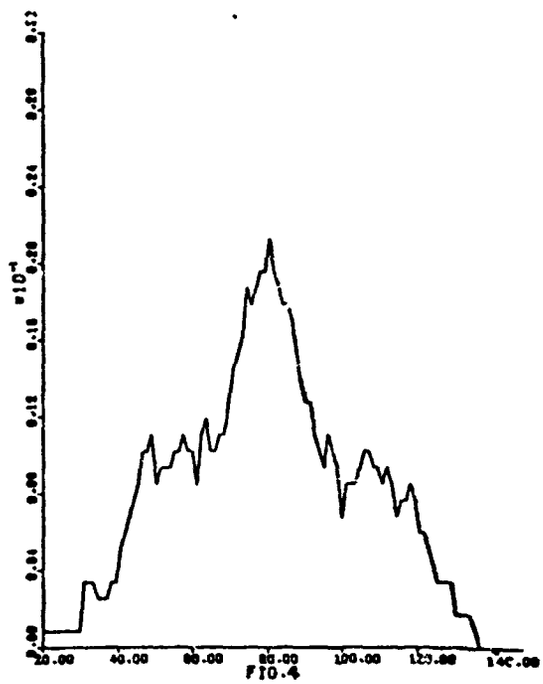


FIG.4

Naive kernel, h = 9.32

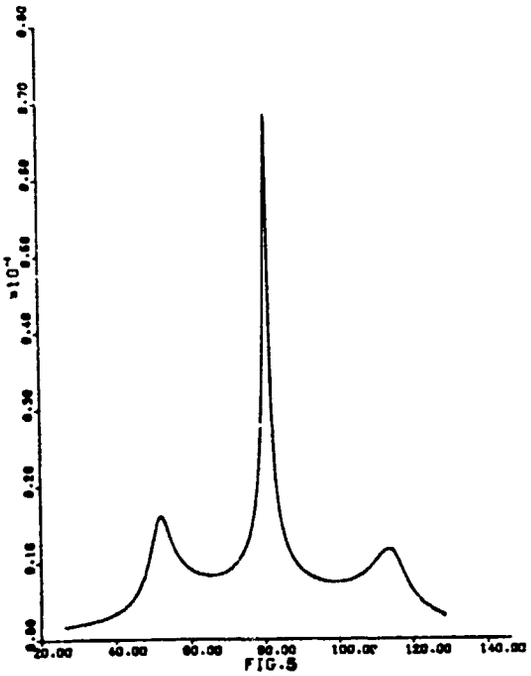


FIG.5

Quantile expansion order 8

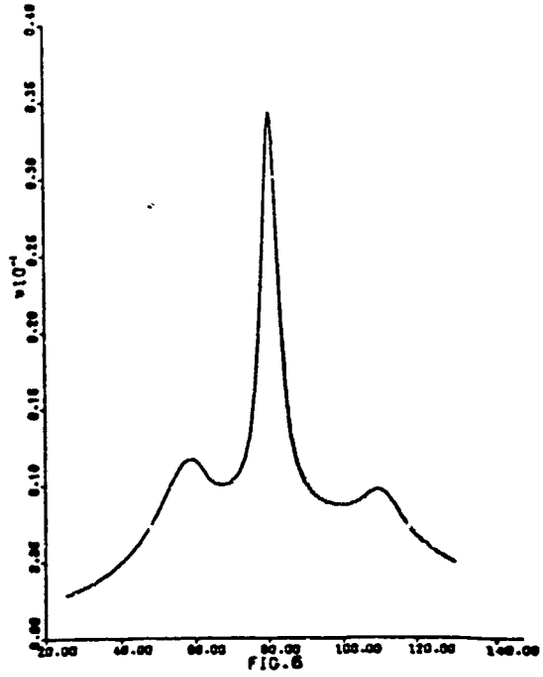


FIG.6

Spline method $\sigma = 1.5$

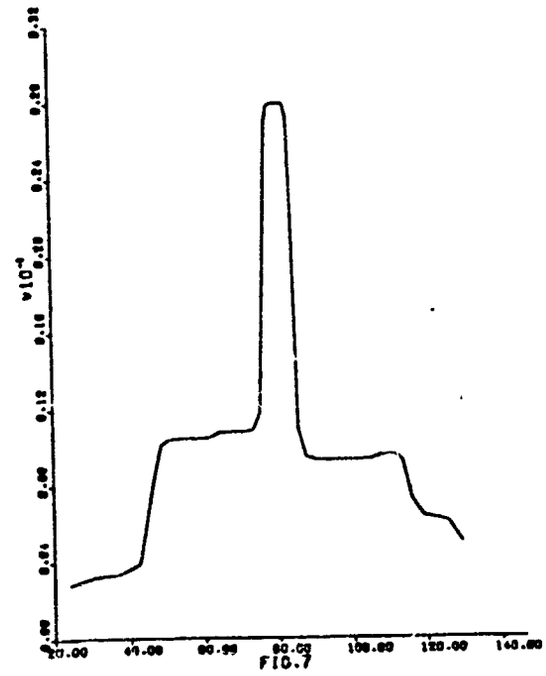


FIG.7

Spline method $\sigma = 15$

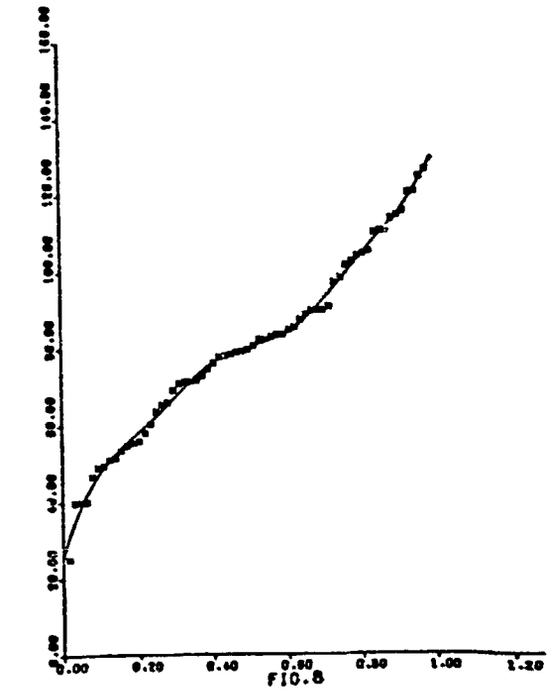


FIG.8

Quantile function obtained from spline $\sigma = 1.5$

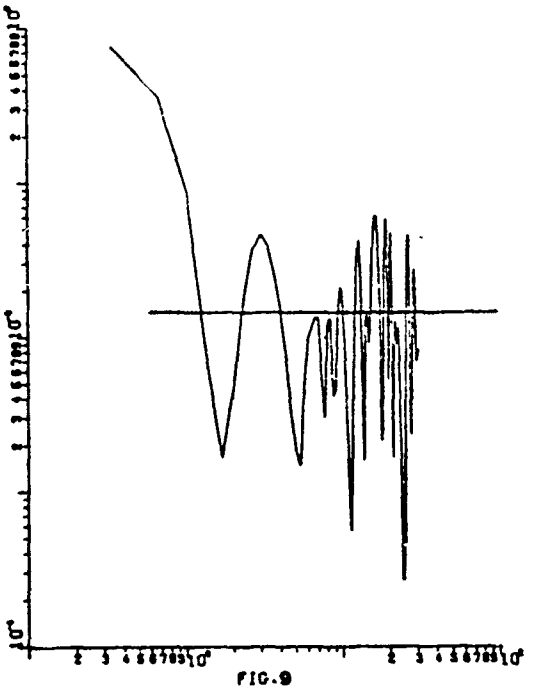


FIG.9

Sample characteristic function
horizontal line at 0.016

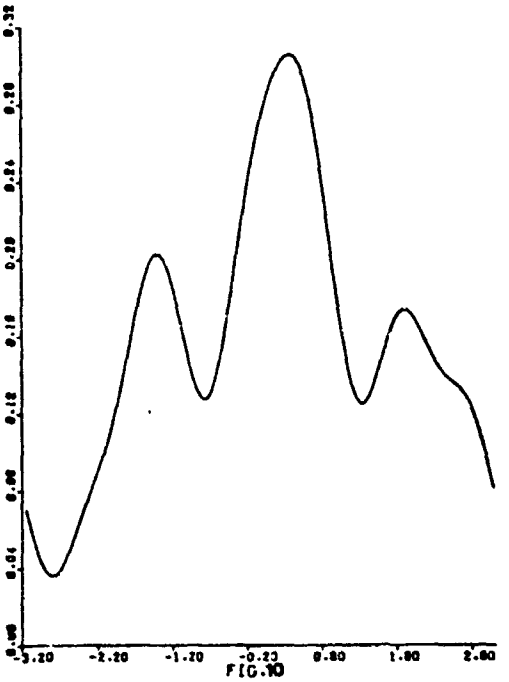


FIG.10

Fourier inversion order 8

2.2.5 The Autoregressive Method.

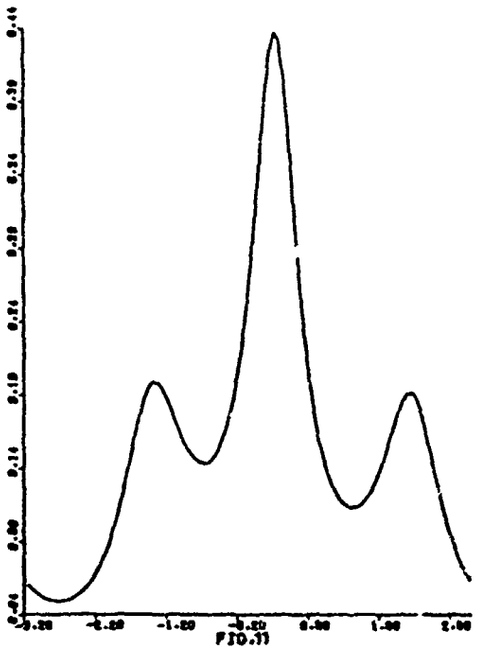
We include several pictures to illustrate the different choices we could make.

In the first four pictures (Fig. 11, 12, 13 and 14), the data filled the interval $[-\pi,\pi]$. These pictures have the same kind of characteristics as Fig. 10 (e.g., the left tail). The number of modes is a non-decreasing function of the order.

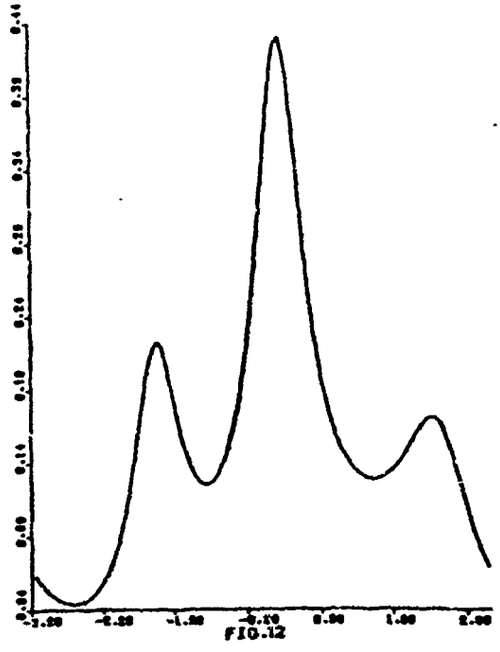
With the data filling only the central 2/3 of the interval $[-\pi,\pi]$, the modes appear much more rapidly (Fig. 15, 16 and 17). The tails are even, but the modes don't stand in the same relation.

What choice should we make?

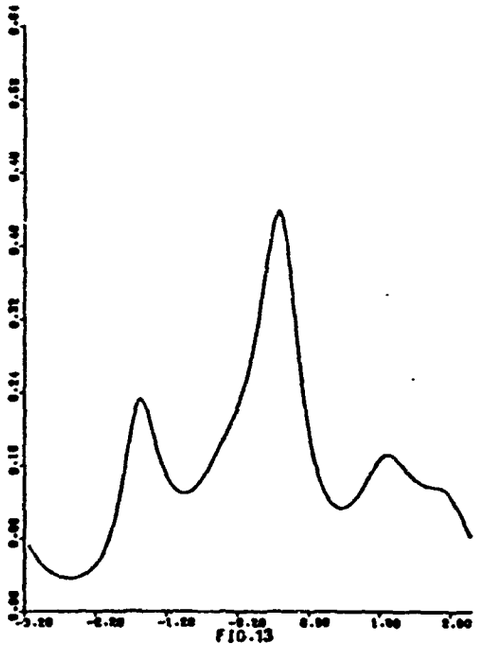
Following one of our conclusions of chapter 1, we can look at the output parameters.



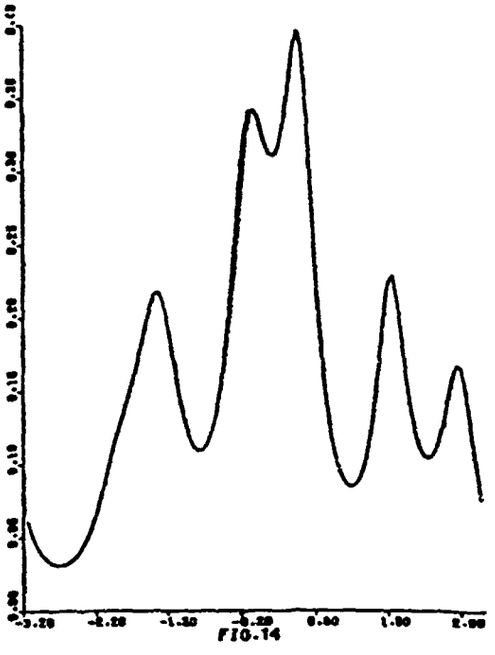
Autoregressive method order 3



Order 4



Order 5



Order 6

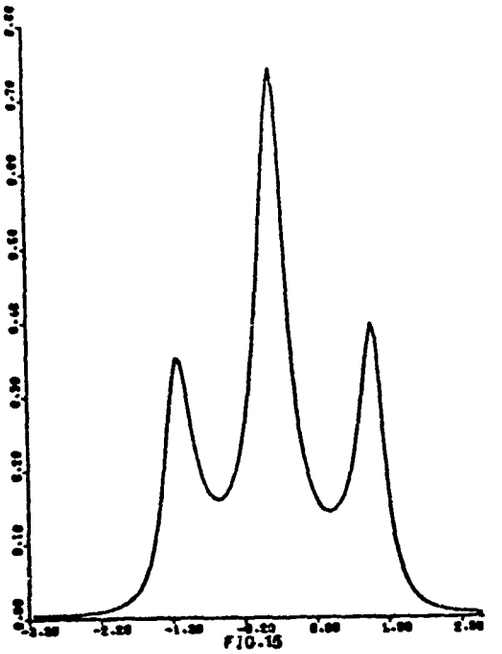


FIG.15

Autoregressive method data rescaled to 66% of $[-\pi,\pi]$ order 3

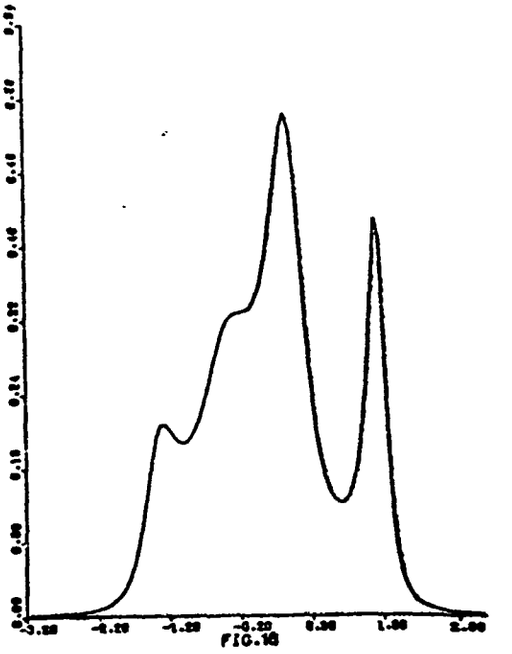


FIG.16

Order 4

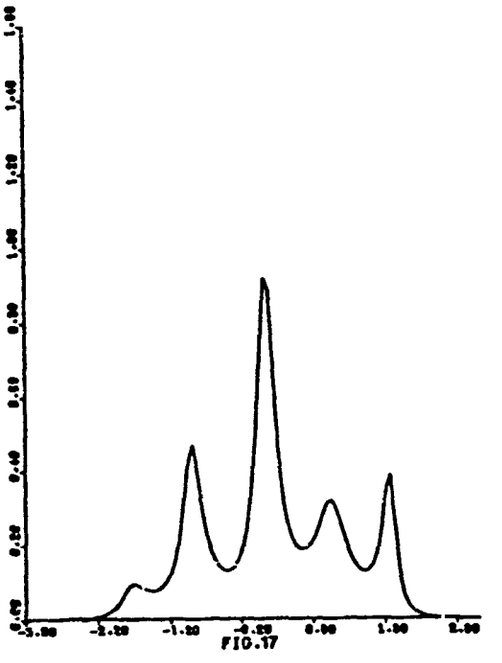


FIG.17

Order 5

Table 2.1 Some parameters of the Buffalo snowfall data density autoregressive estimator (data rescaled to $[-\pi,\pi]$)

Order m	Coefficients $\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	Scale Factor $\hat{K}_m$
1	(-0.2880,0.0792)		0.9107
2	(-0.2943,0.0935)	(0.0075,-0.0516)	0.9082
3	(-0.3064,0.0905)	(0.0546,-0.1084)	0.8564
4	(-0.3094,0.1081)	(0.0554,-0.1174)	0.8517
5	(-0.3160,0.1118)	(0.0487,-0.1422)	0.8427
6	(-0.3221,0.0979)	(0.0609,-0.1319)	0.8242
⋮			
9	(-0.4550,0.1386)	(0.1570,-0.2222)	0.6769

Table 2.2 Some parameters of the Buffalo snowfall data density autoregressive estimator (data rescaled to $\left[-\frac{2\pi}{3}, \frac{2\pi}{3}\right]$)

Order m	Coefficients $\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	Scale Factor $\hat{K}_m$
1	(-0.5843,0.1321)		0.6410
2	(-0.8359,0.2090)	(0.3813,-0.2178)	0.5174
4	(-1.1969,0.2016)	(1.0332,-0.3739)	0.3537
8	(-2.0489,0.2873)	(2.9781,-0.6858)	0.1011

After order 4 in Table 2.1, the parameter K_m decreases by bigger jumps and the coefficients α_{jm} vary more widely from one order to the next.

In Table 2.2, the behavior is the same as in the similar situation we encountered in chapter 1 (Table 1.2). The data did not fill enough of the interval $[-\pi,\pi]$.

Figure 12 seems to be the best choice. To correct the left tail, we could contract the data to 90% of $[-\pi,\pi]$ with minimal effect on the modes.

2.3 Maguire Data

This set of data was studied by Maguire, Pearson and Wynn (1952). It consists of "time intervals in days between explosions in mines involving more than ten men killed, from December 6, 1875 to May 29, 1951." There are 109 data points and only 93 distinct values (see Table B and Fig. B1, B2).

The authors concluded: "none of the tests described in this paper demonstrates lack of homogeneity in the series of time intervals." This set was also studied by Boneva, Kendall and Stefanov (1971) (see Fig. B3).

There were several reasons to consider this set of data, like the possibility of an exponential underlying distribution and the effect of the range.

Table B

Maguire Data

378	36	15	31	215	11	137	4	15	72
96	124	50	120	203	176	55	93	59	315
59	61	1	13	20	189	345	81	286	114
108	188	233	28	22	61	78	99	326	275
54	217	113	32	23	151	361	312	354	58
275	78	17	1205	644	467	871	48	123	457
498	49	131	182	255	195	224	566	390	72
228	271	208	517	1613	54	326	1312	348	745
217	120	275	20	66	291	4	369	338	336
19	329	330	312	171	145	75	364	37	19
156	47	129	1630	29	217	7	18	1357	

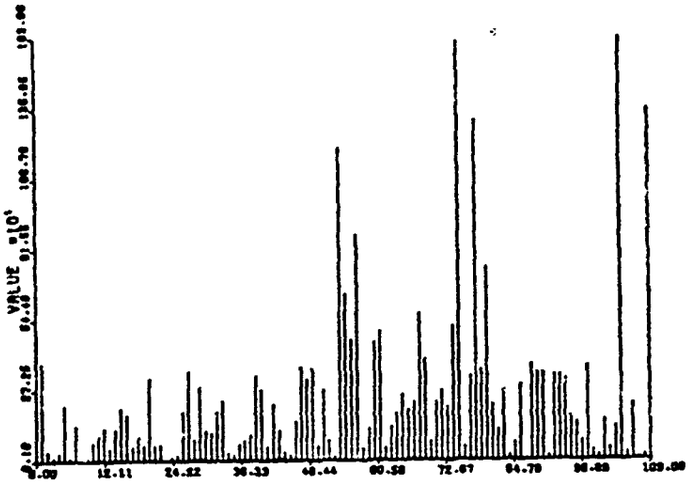


Figure B1
Maguire data

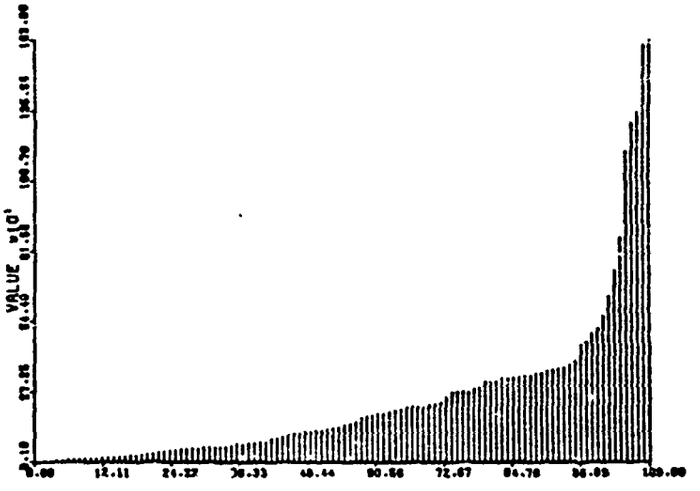


Figure B2
Same data ordered

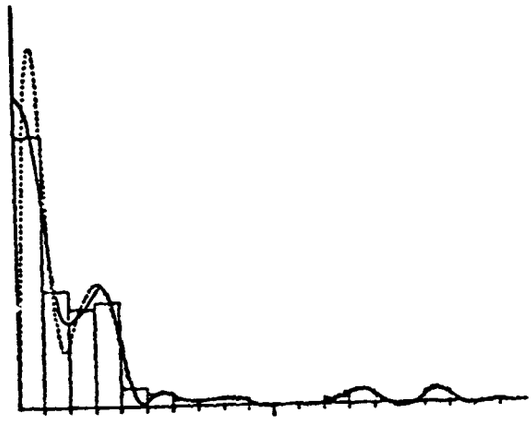


Figure B3

2.3.1 The Kernel Method

The extreme observations have so much effect on the standard deviation of the sample that we can only obtain a very flat estimate of the density using the Parzen kernel. But by omitting the nine largest observations in the computation of the standard deviation, we obtain Figure 18. Despite the "bias" towards normality exhibited by this kernel (as noted in section 2.2.1), we recognize the very short left-hand tail so typical of the exponential density.

2.3.2 The Quantile Expansion Method

It is not possible to get a non-decreasing estimate of the quantile function, and this implies that one cannot form its functional inverse which is necessary to get an estimate of the density.

When we consider only the 93 distinct data points, we get an estimate (Figure 19) which is obviously biased, but nevertheless can be useful in our comparisons.

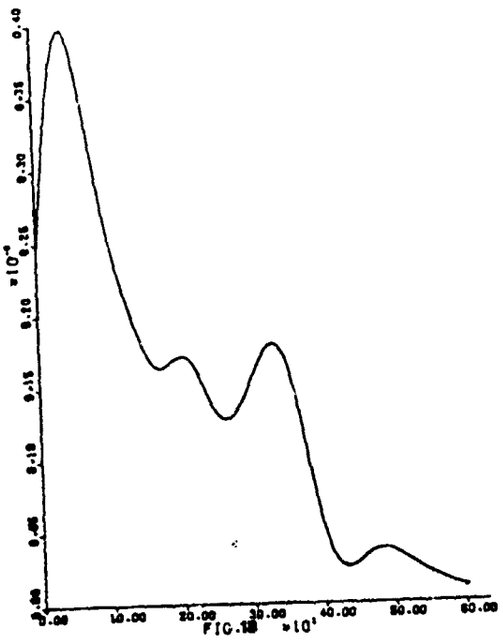
2.3.3 The Spline Method

The spline method proves equally difficult to use on the raw data. Because of the difference of concentration of the data along the real line and the necessity to pack all the data on a same graph, we lose much of our power of resolution. The spline method is also tedious to use as there is no systematic way of positioning the knots.

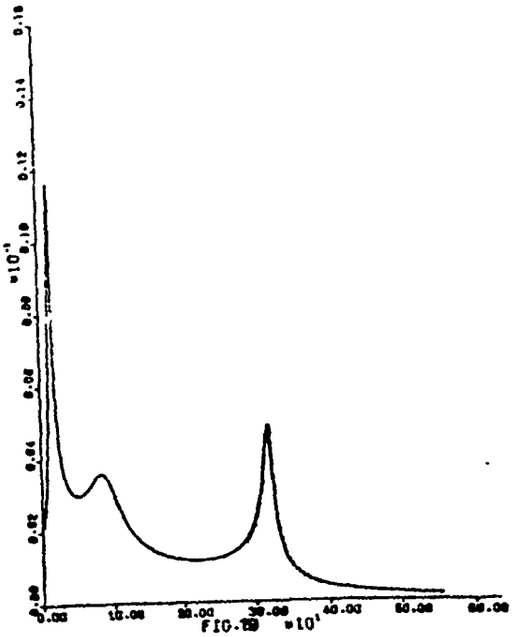
After many trials, we did not produce any satisfactory estimate for mostly the same reason as in the quantile expansion method.

2.3.4 The Weighted Fourier Series Method

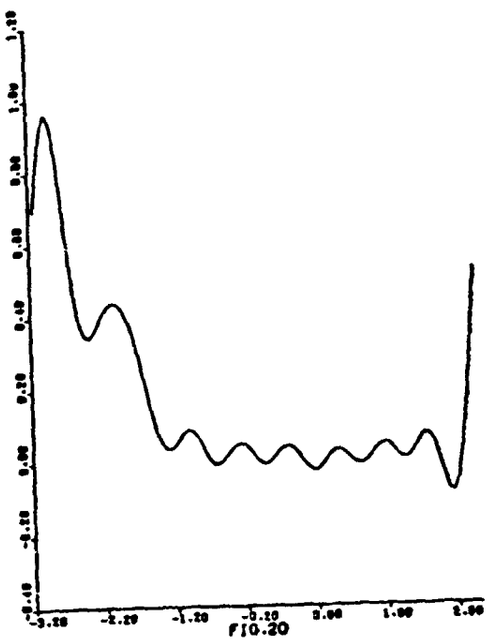
In Figure 20 , we notice that the estimate produced by the weighted Fourier series method has many bumps and is negative at many points. But the two major bumps are located at about the same place as in Figure 18 .



Parzen kernel h = 95.43



Quantile expansion order 8



Fourier inversion order 9

2.3.5 The Autoregressive Method

First we look at the raw data (Figure 21). Notice the large part of the domain where the estimate is essentially zero: this corresponds to an interval that contains only five isolated points. The two modes are at about 50 and 320 as in Figure 18. The coefficients in Table 2.3 seem to indicate order 4 or 5 (Fig. 21 or 22).

Table 2.3 Some parameters of the Maguire data density autoregressive estimator

Order	Coefficients		Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{K}_m$
3	(0.7417,-0.7200)	(-0.0927,-0.2868)	0.3120
4	(0.7149,-0.7027)	(-0.0288,-0.3606)	0.2792
5	(0.6821,-0.8211)	(-0.1105,-0.3999)	0.2391
6	(0.6805,-0.9146)	(-0.2540,-0.4430)	0.2245
⋮			
13	(1.2903,-2.1401)	(-1.2134,-3.0118)	0.0152

In the exponential case, we have found that symmetrization yields good results. In the real data case all we need to do to symmetrize is to use the real part of the Fourier transform and evaluate the density on $[0,\pi]$. We notice again that for the same order the symmetrized estimate is smoother (Fig. 23 vs Fig. 22). This allows us to consider higher orders (Fig. 25).

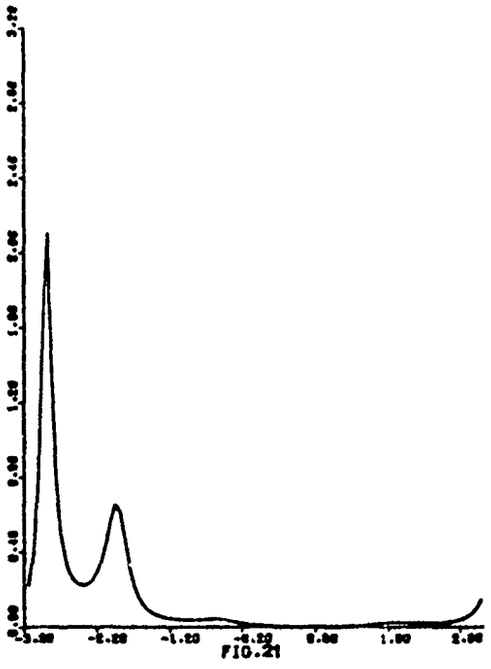


FIG.21

Autoregressive method order 4

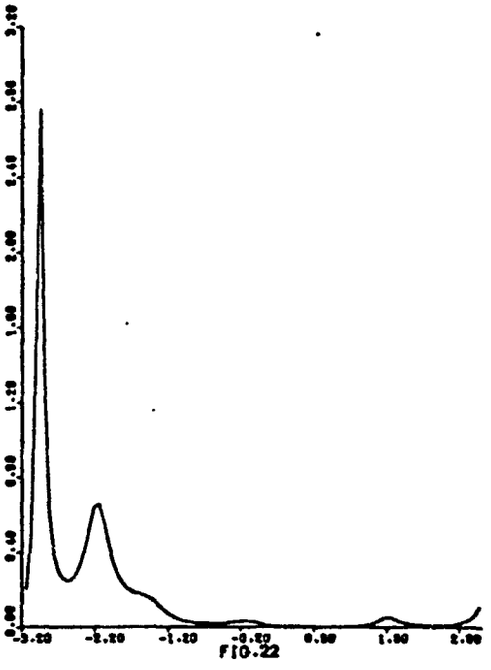
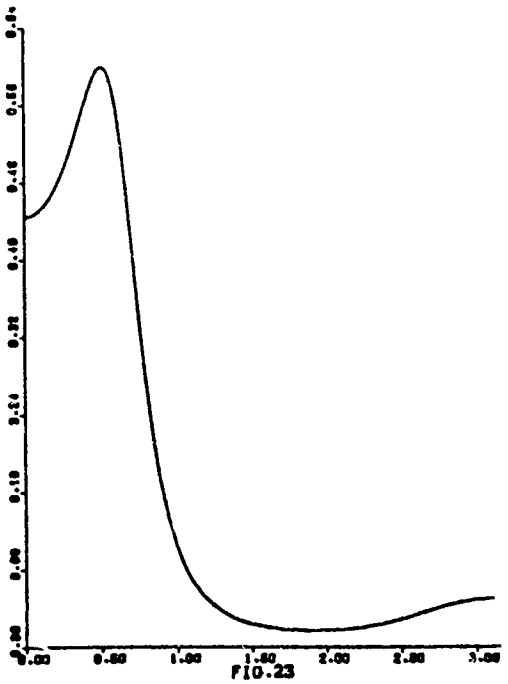
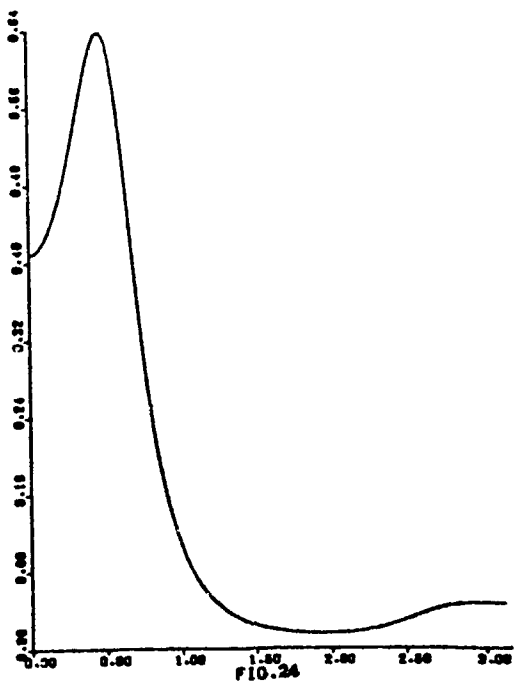


FIG.22

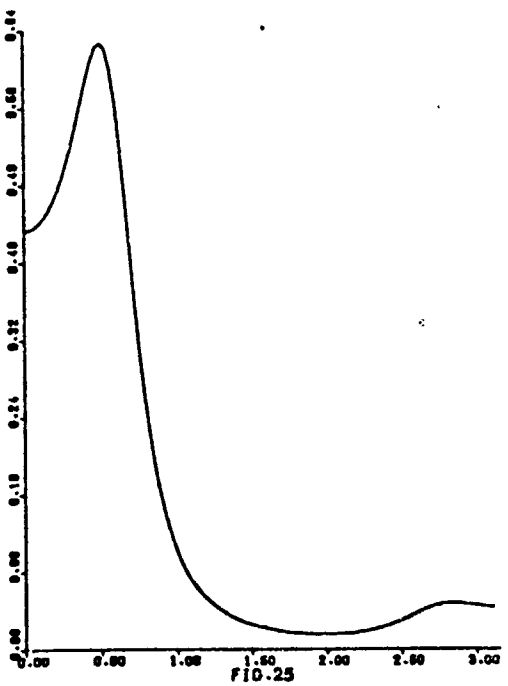
Order 5



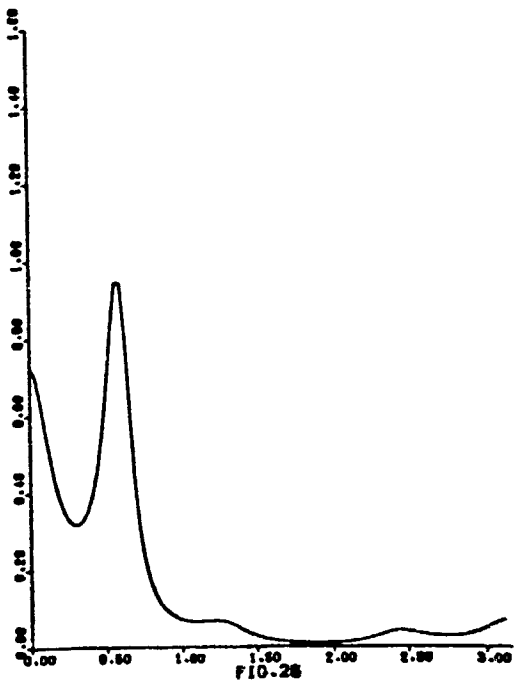
Symmetrized data order 5



Order 6



Order 7



Order 8

Table 2.4 Some parameters of the Maguire data density autoregressive estimator (symmetrized)

Order	Coefficients				Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{\alpha}_{4m}$	$\hat{\alpha}_{5m}$	$\hat{K}_m$
5	-.7156	-.2775	.0539	-.0764	.4251
6	-.7188	-.2753	.0424	-.1061	.4244
7	-.7200	-.2723	.0303	-.0982	.4240
8	-.7128	-.2879	.0227	-.2039	.3973
9	-.7826	-.2456	-.0340	-.1976	.3666
10	-.8548	-.1290	-.1001	-.1463	.3419

We notice that the shape changes slowly (Fig. 23, 24, 25 and 26) from one order to the next as the parameters remain very stable. But from the big change in the parameters of orders 8 and 9, it would seem that order 8 is indicated (Fig. 25).

We then proceed to the square-root transformation. The autoregressive estimates are much smoother than the one obtained from the weighted Fourier series method (Fig. 27-29 compared to Fig. 20). There are no bumps in the right-hand tail, which is an improvement over Figures 24-26. By looking at Table 2.5, we can narrow our choice among orders 2 to 5. The shape is pretty much the same, but the location of the modes is moved around.

Table 2.5 Some parameters of the Maguire data density autoregressive estimator (square root transformation)

Order	Coefficients		Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{K}_m$
1	(-0.0376,-0.5504)		0.6956
2	(-0.0751,-0.7748)	(-0.4011,0.0954)	0.5773
3	(-0.0758,-0.8066)	(-0.4605,0.1050)	0.5735
4	(-0.0784,-0.8073)	(-0.4644,0.1214)	0.57[illegible]
5	(-0.0750,-0.8092)	(-0.4743,0.1206)	0.5662
7	(-0.1020,-0.8337)	(-0.4989,0.1606)	0.5361

Figures 21, 25 and 29 have their second mode at about 320 , like the kernel estimate in Figure 18 and the quantile estimate of Figure 19.

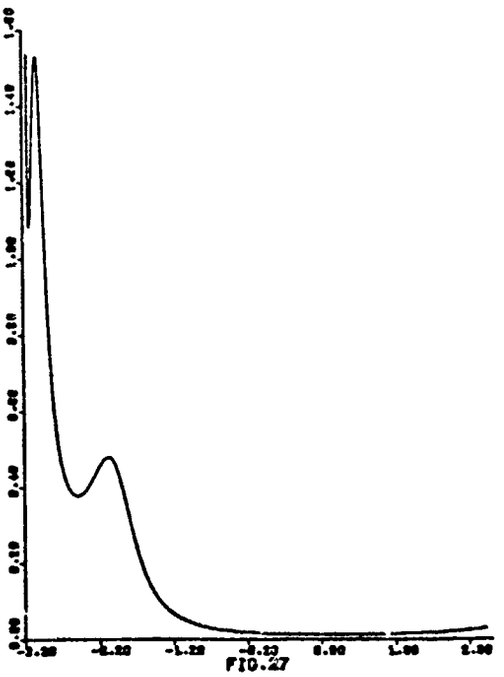


FIG.27

Square-root transformation
order 2

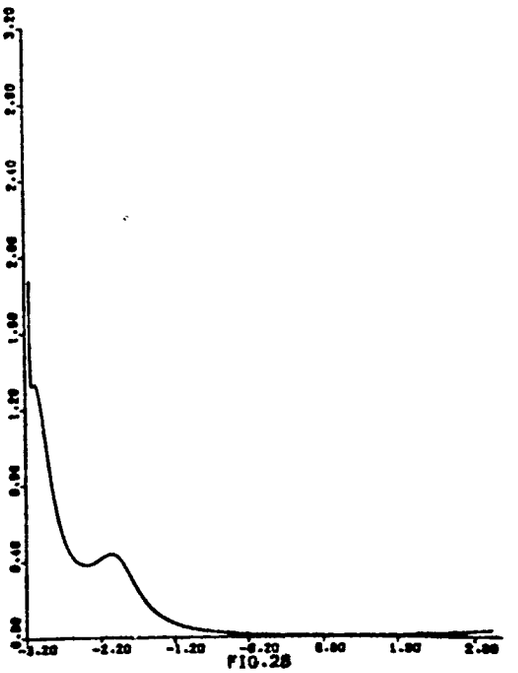


FIG.28

Order 4

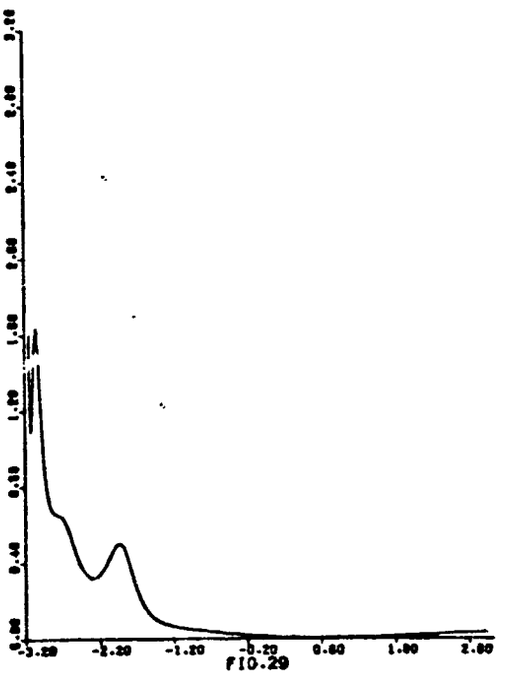


FIG.29

Order 5

2.4 Bliss Data

This set of data is taken from Bliss (1967) Table 7.1 . It consists of a 28-cell histogram for the lengths of survival in days of 1110 mice inoculated uniformly with malaria. This set was also used by Boneva, Kendall and Stefanov (1971) (see Fig. C1).

We use it here to compare the behavior of the different methods with respect to grouped data, i.e., when smoothing a histogram.

Table C

Bliss Data

Midpoint	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5
Frequency	25	90	75	69	48	36	29	30	33	44
Midpoint	14.5	15.5	16.5	17.5	18.5	19.5	20.5	21.5	22.5	23.5
Frequency	29	40	51	51	71	65	78	75	48	30
Midpoint	24.5	25.5	26.5	27.5	28.5	29.5	30.5	31.5		
Frequency	35	17	15	13	4	6	2	1		

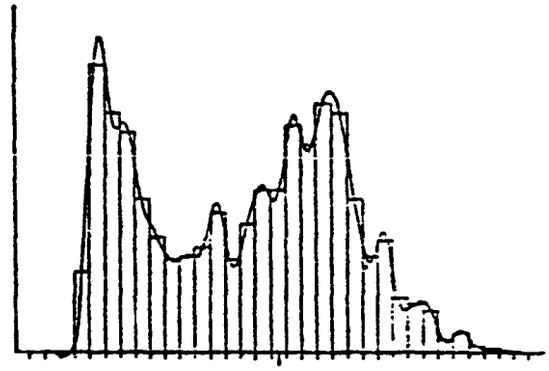


Figure C1

2.4.1 The Kernel Method

The kernel method produces once again a smooth estimate (Fig. 30), with two modes at 7 and 20. The modes stand in inverse relation compared to Figure C1. For the used h, there are less points contributing to the estimation around the first mode than around the second. A smaller value of h would give the proper relation, but it would also give spurious bumps.

2.4.2 The Quantile Expansion Method

The quantile method cannot be used with grouped data unless one unravels the histogram by distributing the frequency count of a cell over its width or by assigning to the midpoint a multiplicity equal to the frequency count. Using the latter we obtain Figure 31.

2.4.3 The Spline Method

In Figure 32, there is an artificial mode at 4, outside the observed range, created by the initial conditions that have to be imposed. The mode at 7 is still present though obscured. By changing the first derivative at the origin of the quantile function from 50 to 22, we eliminate the artificial mode (Fig. 33) without effect on the second mode.

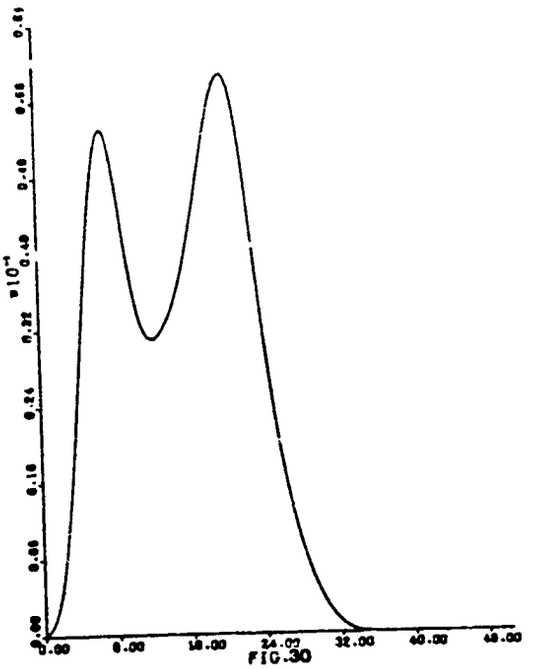


FIG.30

Parzen kernel h = 6.04

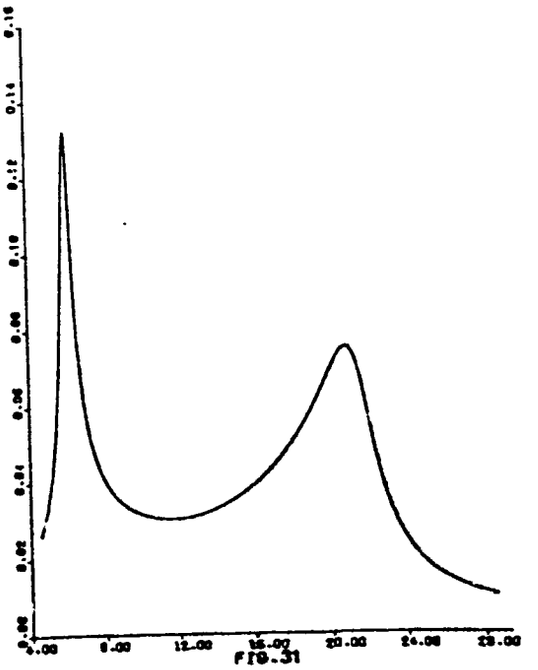


FIG.31

Quantile expansion order 8

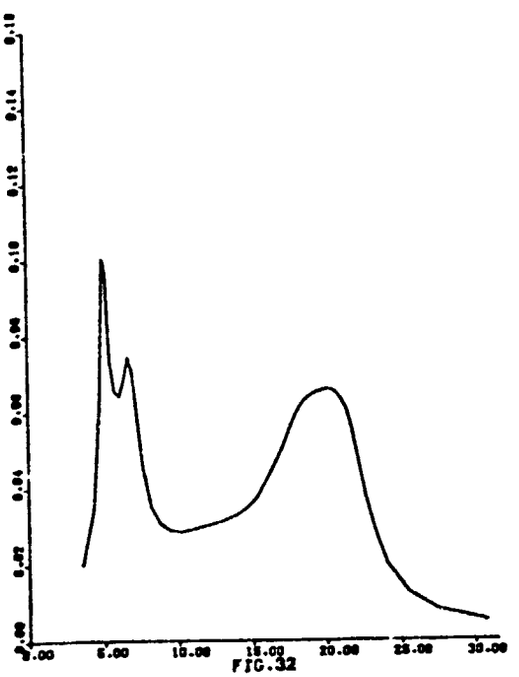


FIG.32

Spline SLP1 = 50

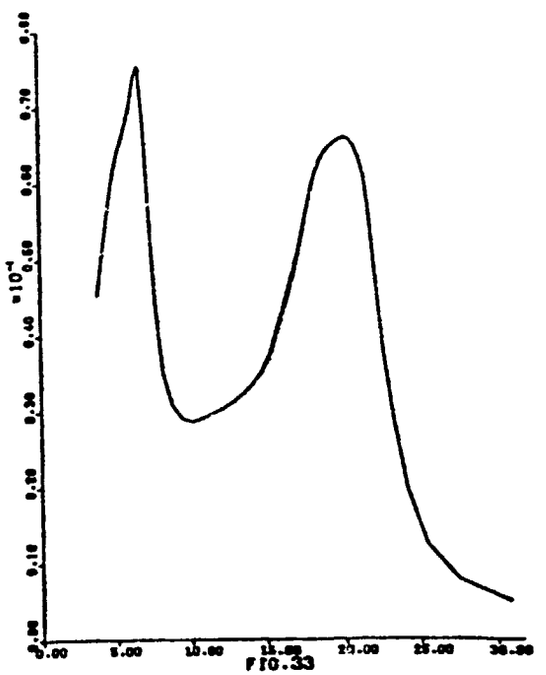


FIG.33

Spline SLP1 = 22

2.4.4 The Weighted Fourier Series Method

There is no difficulty to handle grouped data in the weighted Fourier series method. But there is always the same problem with the tails (Figure 34).

2.4.5 The Autoregressive Method

Not to include any zero cell at either end is equivalent to having the data fill the interval $[-\pi,\pi]$. From Table 2.6 and the previous pictures, the estimates of order 2 and 3 seem reasonable (Fig. 35 and 36).

If we include a zero cell at each end, the data points are contracted to fill only 93% of $[-\pi,\pi]$; the pictures differ very slightly (Fig. 37 and 38), except at the tails as expected.

Table 2.6 Some parameters of the Bliss data density autoregressive estimator

Order	Coefficients		Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{K}_m$
1	(-0.1437,-0.0767)		0.9734
2	(-0.1484,-0.1311)	(-0.1316,0.3082)	0.8640
3	(-0.2170,-0.1535)	(-0.1314,0.3508)	0.8240
4	(-0.2482,-0.1578)	(-0.0832,0.3769)	0.8064
5	(-0.2629,-0.1641)	(-0.0607,0.3927)	0.7967

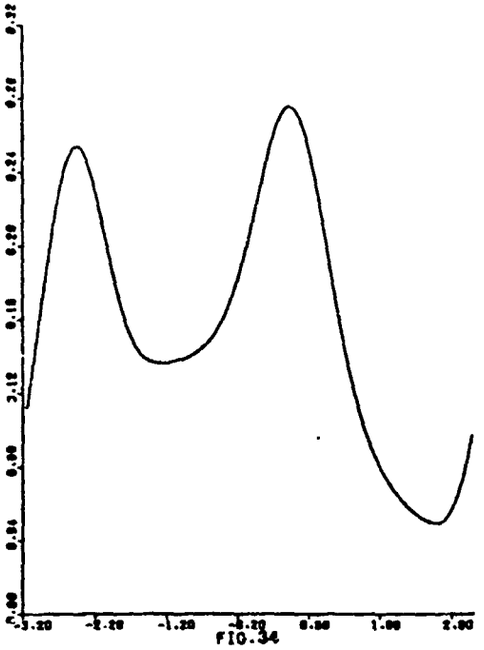


FIG.34

Fourier inversion order 6

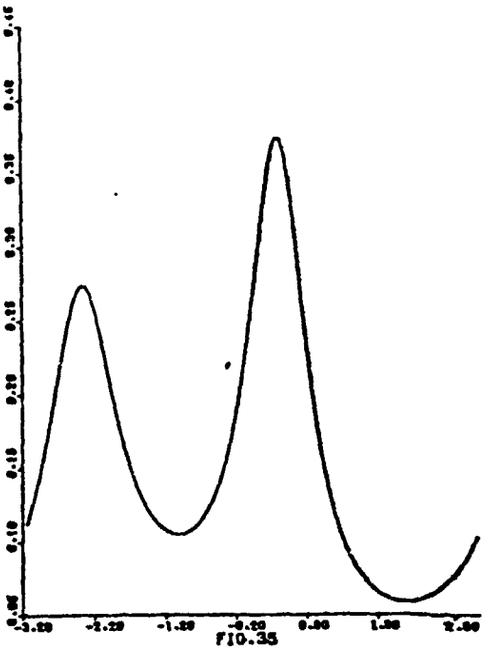


FIG.35

Autoregressive method order 2

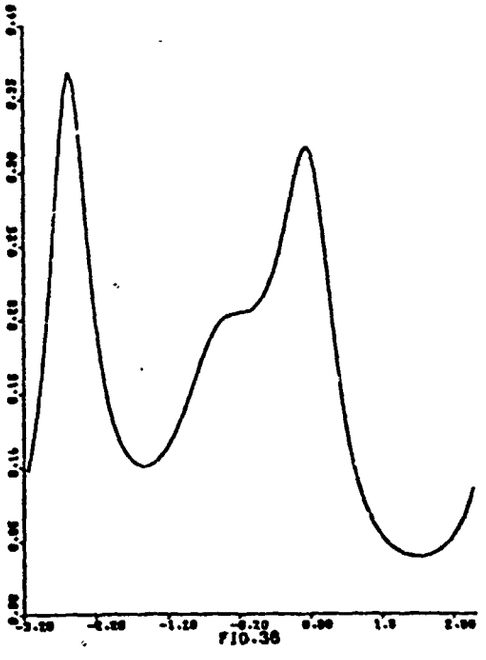


FIG.36

Autoregressive method order 3

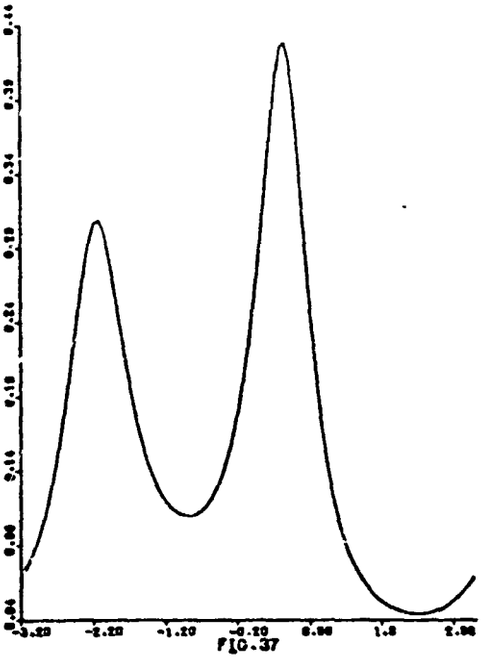


FIG.37

Autoregressive method order 2,
with zero cell added at each end

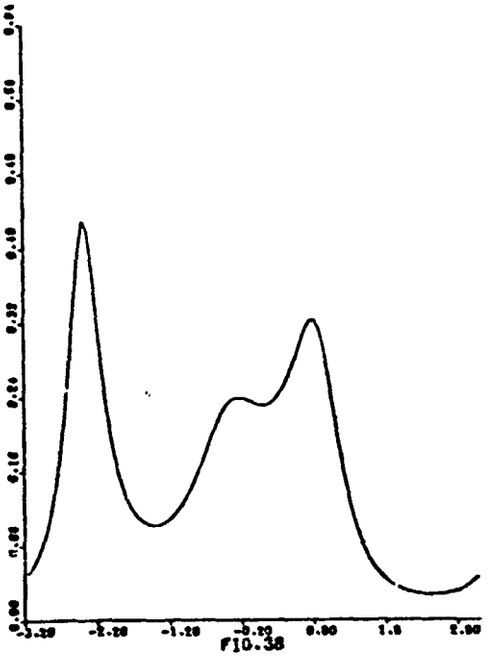


FIG.38

Order 3

2.5 Hazard Estimation

The Maguire data set can also be studied from the point of view of hazard estimation. It is difficult to judge whether the estimates produced by the autoregressive method are reasonable. But as there is a one-to-one relationship between the hazard function and the density function, we can make the comparison in terms of the density.

Our best estimate of the density was Figure 21 which gives the hazard function on Figure 39, obtained by the indirect estimation procedure

$$\hat{h}(x) = \frac{f_m(x)}{1 - \int_0^x f_m(u)\, du}$$

Table 2.7 Some parameters of the Maguire data hazard function autoregressive estimator

Order	Coefficients		Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{K}_m$
1	(0.4214,-0.2538)		3.5558
2	(0.4873,-0.3144)	(0.0511,-0.1748)	3.4378
3	(0.5085,-0.3099)	(0.1114,-0.2080)	3.3893
4	(0.4864,-0.3551)	(0.0677,-0.2978)	2.7822
5	(0.4262,-0.4846)	(-0.0282,-0 3139)	2.7696

A look at Table 2.7 would leave the choice to be made among orders 2, 3 and 4 (Fig. 40, 41 and 42). From these hazard est mates, we can evaluate the density indirectly by

$$\hat{f}(x) = h_m(x) \exp(- \int_0^x h_m(u)\, du)$$

By this process, we obtain Figures 43, 44 and 45.

As Figure 42 is closest to Figure 39, so is Figure 45 to Figure 21.

We note that the direct estimation of the hazard function is difficult at the right-hand tail where it becomes infinite. Also, in the indirect processes, we have to perform some numerical integration: we used the trapezoidal rule which should be adequate as long as the function we integrate does not have too many sharp teeth.

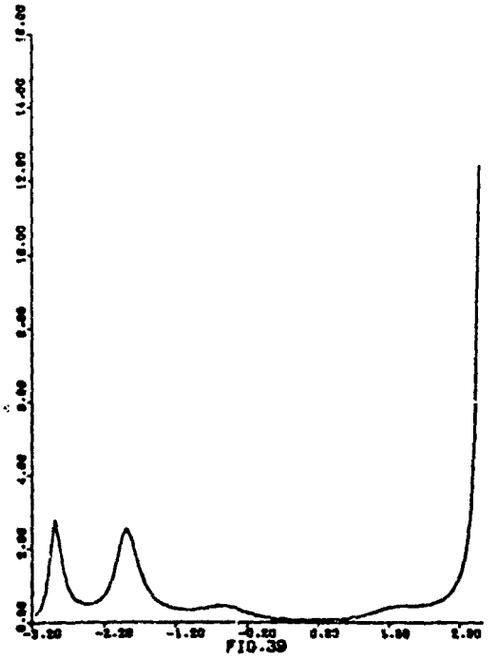


FIG.39

Hazard obtained from density
order 4

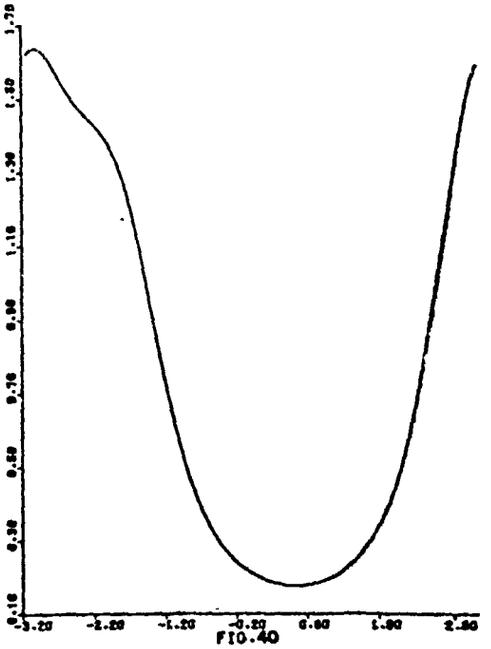


FIG.40

Hazard obtained directly
order 2

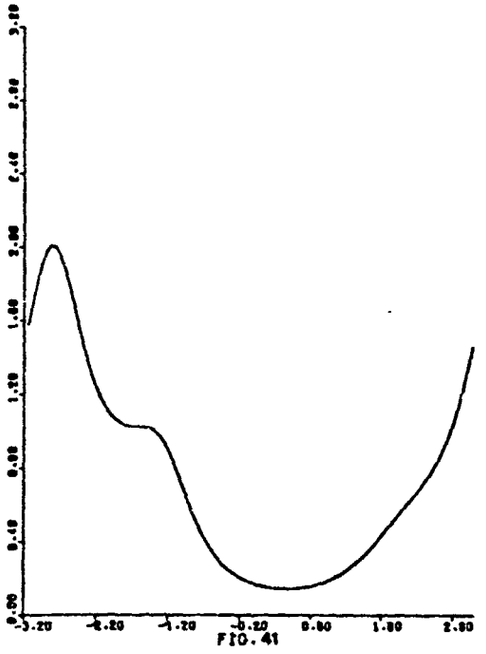


FIG.41

Order 3

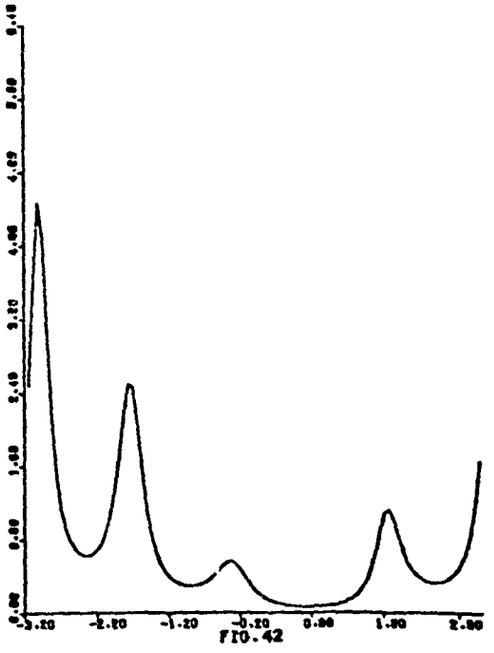


FIG.42

Order 4

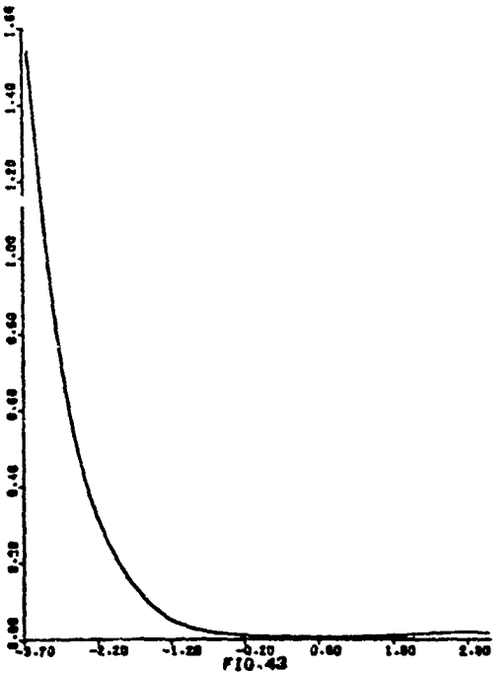


FIG.43

Density obtained from hazard order 2

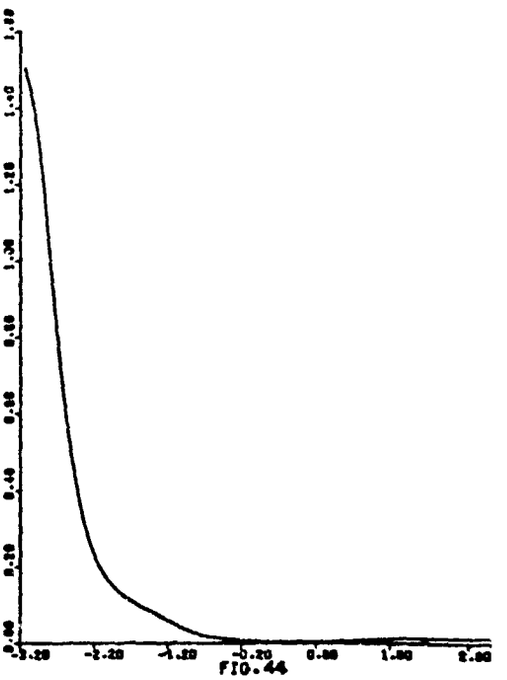


FIG.44

Order 3

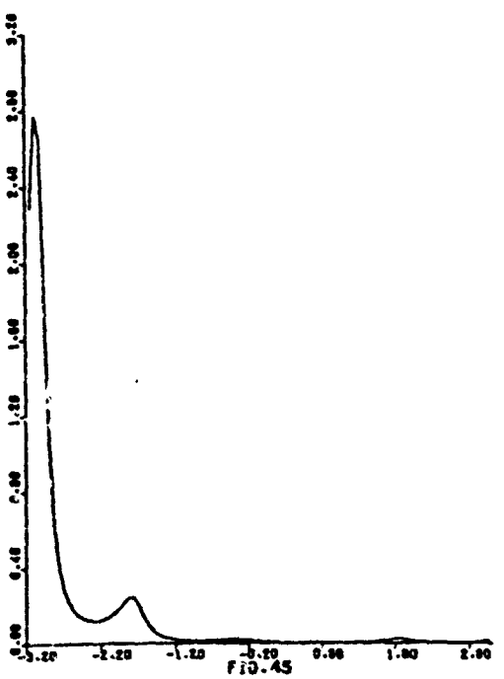


FIG.45

Order 4

2.6 Estimating the Density via the Quantile Function

There is still another route open to estimate the density function: via the quantile function.

We use the Fourier transform of the sample quantile function $Q_n(t)$ to get the autoregressive estimate of order m of the sparsity function, $\hat{q}_m(t)$. We integrate $\hat{q}_m(t)$ to get $\hat{Q}_m(t)$.

Now using the relation given in (0.A.1.2), the density estimator is

$$f(\hat{Q}_m(t)) = \frac{1}{\hat{q}_m(t)}$$

We illustrate it only on the Maguire data, Figures 46 to 49. The general shape is well preserved (compare with Fig. 21-22). Table 2.8 lists some of the parameters. Orders 2, 3 or 4 seem likely candidates.

Table 2.8 Some parameters of the Maguire data sparsity function autoregressive estimator

Order	Coefficients		Scale Factor
m	$\hat{\alpha}_{1m}$	$\hat{\alpha}_{2m}$	$\hat{K}_m$
1	(0.6615,0.2498)		814.9191
2	(0.3546,0.3447)	(-0.4534,-0.0277)	646.7340
3	(0.2663,0.3703)	(-0.3986,-0.1115)	620.2509
4	(0.2543,0.3741)	(-0.3808,-0.1301)	617.8502
5	(0.2505,0.3790)	(-0.3766,-0.1491)	611.7474
6	(0.2254,0.3736)	(-0.3687,-0.1701)	571.0632

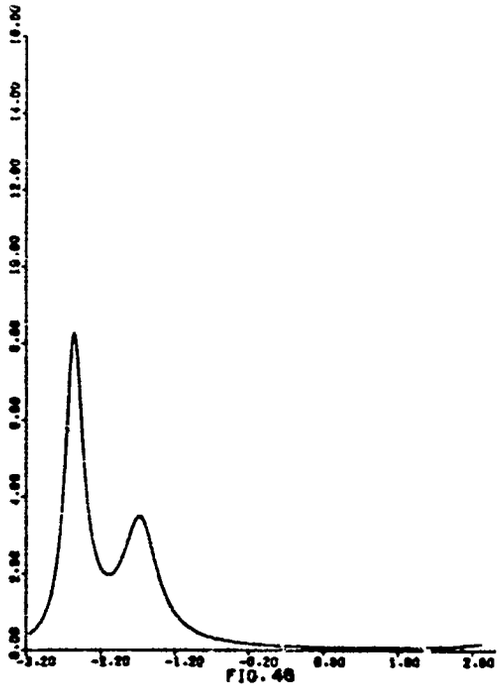


FIG. 46

Density obtained from sparsity order 2

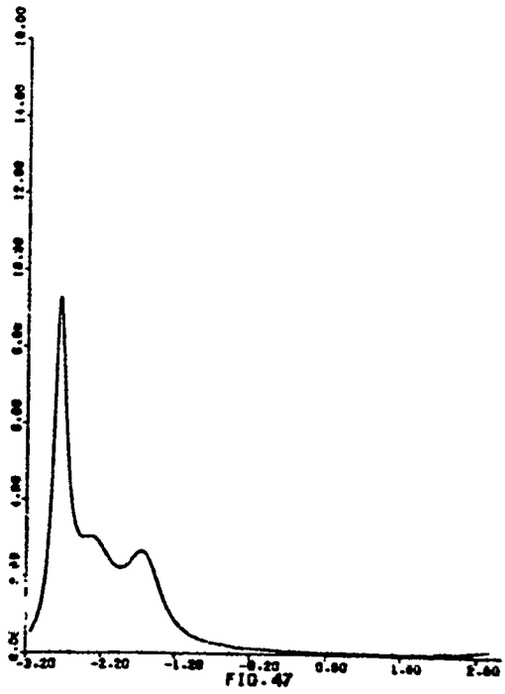


FIG. 47

Order 3

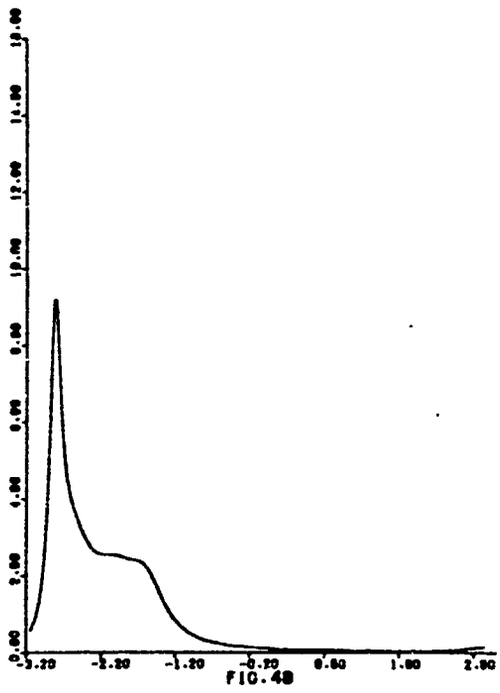


FIG. 48

Order 4

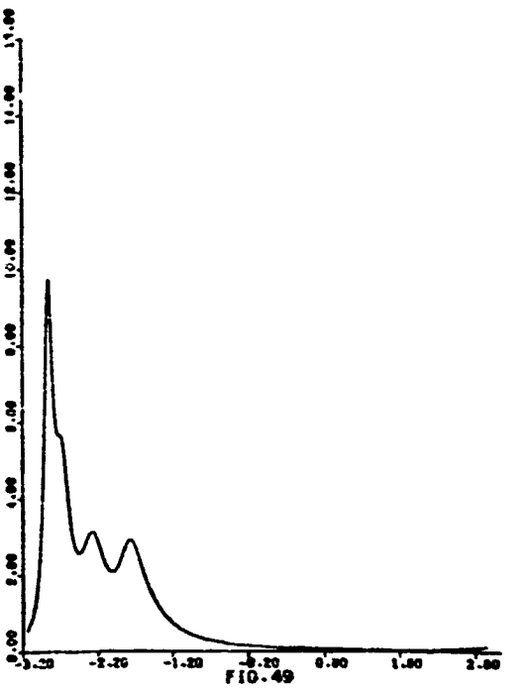


FIG. 49

Order 5

2.7 Conclusions

1. We can produce with ease, using the autoregressive method, estimates that compare favorably with those of the competing techniques. We never get negative estimates as is possible with the quantile method, the spline method or the weighted Fourier series method.

2. The spline method can be very tedious to use when problems with the knot points arise.

3. The kernel method performs very well, but the shape of the kernel has an influence on the shape of the estimate.

4. In the autoregressive method, the data should fill at least 90% of the interval $[-\pi,\pi]$.

5. Picking the best order is made easier by looking at the output parameters. This is another advantage over the weighted Fourier series method. The quantile method also produces output parameters, the $\hat{\lambda}_j$'s , mentioned in our Introduction. Sillitto (1969) gives interpretation to some of them.

6. A stricter rule to pick the best order of the autoregressive estimator would be preferable.

7. It seems that transformations of the data can improve the properties of the autoregressive estimators, notably in the tails.

APPENDIX

2.A.1 Sample Programs for Estimation

We include in this appendix three sample programs, one for each of the approaches we used.

Each program is divided in 3 parts:

I - Computing the estimated R(•) sequence

II - Solving the Yule-Walker equations in AUTOREG (see appendix to Chapter 3).

III - Computing the estimate of the density.

1st program: Estimating the density of Maguire data.

In Part I, we perform the square-root transformation of the data Y(•) and compute the estimated R(•) sequence in GCSPHI, stored in A(•). FREQ(•) contains the frequency of each data point.

In Part II, we solve the Yule-Walker equations. F(•) contains the estimated density of the square root.

In Part III, we transform back to the original scale.

Subroutine GCSPHI is equivalent to FOURSTI except that it also computes the square modulus of the R(•) sequence stored in PHI2. This feature is not needed in the autoregressive method.

2nd program: Estimating the hazard of Maguire data.

In Part I, CF(•) contains the empirical c.d.f. and F(•)

contains the first difference of the estimated integrated hazard

$\log\left(1 - \frac{n}{n+1} F_n(\cdot)\right)$.

Part II is as before.

In Part III, we transform back from the hazard to the density.

3rd program: Estimating the quantile of Maguire data.

In Part I, FREQ(•) is the frequency of each data point, CF(•) is the empirical distribution function. Then, FREQ(•) is modified to become the first difference of the data Y(•) , i.e., the first difference of the empirical quantile function.

The two other parts are as before.

In these programs, we always rescale the data to $[-\pi,\pi]$ using subroutines CENTER and XSCALE .

```
      PROGRAM TENSE(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      DIMENSION FREQ(93)
      COMPLEX A(2?),ALPHA(2?),PHI(2?),JH,KH
      DIMENSION R(1?),C(2?)
      DIMENSION X(1?2),F(1?2)
      DIMENSION Y(1??
      DIMENSION G(1?2)
      PI=4.*ATAN(1.?)
      TWOPI=2.*PI
      M=7
      M=M-1
      L=1
      NI=93
      N=NI
      READ(5,1??) (Y(I),I=1,N)
1?? FORMAT(8F1?.?)
      READ(5,1?1) (FREQ(I),I=1,NI)
1?1 FORMAT(8F1?.?)
      DO 4 I=1,NI
      Y(I)=SQRT(Y(I))
      FREQ(I)=FREQ(I)/199.
    4 CONTINUE
      XMID=?.5*(Y(1)+Y(N))
      XRNG=Y(N)-Y(1)
      SCL=TWOPI/XRNG
      CALL CENTER(Y,Y,N,XMID)
      CALL XSCALE(Y,Y,N,SCL)
      CALL GOSPHI(Y,N,R,C,PHI,FREQ,M,L)
      A(1)=CMPLX(1.,?.)
      DO 15 J=2,M
      A(J)=CMPLX(R(J-1),C(J-1))
   15 CONTINUE
      K=1??
      FK=K
      NP=K
      DO 1 J=2,M
      DO 13 I=1,K
      X(I)=-PI+(I-1)*TWOPI/FK
   13 CONTINUE
      CALL AUTOREG(A,J,M,K,ALPHA,PHI,JH,KH,X,F)

      CALL XSCALE(X,X,-K,SCL)
      CALL CENTER(X,X,-K,XMID)
      CALL XSCALE(F,F,K,SCL)
      DO 24 I=1,NP
      F(I)=?.5*F(I)/X(I)
      X(I)=X(I)*X(I)
   24 CONTINUE
    1 CONTINUE

      END
```

I

II

III

```
      SUBROUTINE GOSPHI(X,N,CPHI,SPHI,PHI2,FREQ,M,L)

      DIMENSION X(N),CPHI(1),SPHI(1),PHI2(1)
      COMPLEX PHI2
      DIMENSION FREQ(N)
      PI=4.*ATAN(1.0)
      EN=FLOAT(N)
      EL=FLOAT(L)
      ML=M*L
      DO 50 J=1,ML
      CPHI(J)=0.
      SPHI(J)=0.
   50 CONTINUE
      DO 100 I=1,N
      IF(ABS(X(I)).GT.PI) GO TO 104
      XX=X(I)/EL
      C1=COS(XX)
      C0=2.*C1
      C2=1.
      S1=SIN(XX)
      S2=0.
      CPHI(1)=CPHI(1)+C1*FREQ(I)
      SPHI(1)=SPHI(1)+S1*FREQ(I)
      IF(ML-1) 101,101,102
  102 CONTINUE
      DO 450 J=2,ML
      CPHI(J)=CPHI(J)+CSREC(C0,C1,C2)*FREQ(I)
      SPHI(J)=SPHI(J)+CSREC(C0,S1,S2)*FREQ(I)
  450 CONTINUE
  101 CONTINUE
  100 CONTINUE
      DO 200 J=1,ML
      C=CPHI(J)
      S=SPHI(J)
      PHI2(J)=C*C+S*S
  200 CONTINUE
      GO TO 5
  104 WRITE(6,103)
  103 FORMAT(1H1,2X,24HDATA NOT SCALED PROPERLY)
    5 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CENTER(X,Y,N,XMID)
      DIMENSION Y(1)
      DIMENSION X(1)
      XPT=XMID
      IF(N.GT.0) XPT=-XMID
      NN=IABS(N)
      DO 9 I=1,NN
    9 Y(I)=X(I)+XPT
      RETURN
      END


      SUBROUTINE XSCALE(X,Y,N,SCL)
      DIMENSION Y(1)
      DIMENSION X(1)
      PTSCL=SCL
      IF(N.LT.0) PTSCL=1./PTSCL
      NN=IABS(N)
      DO 10 I=1,NN
   10 Y(I)=X(I)*PTSCL
      RETURN
      END
```

```
      PROGRAM TENSE(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      DIMENSION FREQ(93)
      COMPLEX A(25),ALPHA(25),PHI(25),JH,KH
      DIMENSION R(100),C(100)
      DIMENSION X(1,2),F(102)
      DIMENSION Y(100)
      DIMENSION CF(1,0),G(100)
      PI=4.*ATAN(1.0)
      TWOPI=2.*PI
      M=5
      L=1
      NI=93
      N=NI
      READ(5,100) (Y(I),I=1,N)
  100 FORMAT(8F10.0)
      READ(5,101) (FREQ(I),I=1,NI)
  101 FORMAT(8F10.3)
      CF(1)=FREQ(1)/110.
      F(1)=-ALOG(1.-CF(1))
      DO 8 I=2,NI
      CF(I)=CF(I-1)+FREQ(I)/110.
      F(I)=-ALOG(1.-CF(I))+ALOG(1.-CF(I-1))
    8 CONTINUE
      XMID=.5*(Y(1)+Y(N))
      XRNG=Y(N)-Y(1)
      SCL=TWOPI/XRNG
      CALL CENTER(Y,Y,N,XMID)
      CALL XSCALE(Y,Y,N,SCL)
      CALL GOSPHI(Y,N,R,C,PHI,F,M,L)
      A(1)=CMPLX(-ALOG(1.-CF(NI))-F(1),0.)
      DO 15 J=2,M
      A(J)=CMPLX(R(J-1),C(J-1))
   15 CONTINUE
      K=100
      FK=K
      NP=K
      FNP=NP
      H=PI/FNP
      DO 1 J=2,M
      DO 13 I=1,K
      X(I)=-PI+(I-1)*TWOPI/FK
   13 CONTINUE
      CALL AUTOREG(A,J,M,K,ALPHA,PHI,JH,KH,X,F)
      G(1)=F(1)*H
      DO 24 I=2,NP
      G(I)=G(I-1)+H*(F(I-1)+F(I))
   24 CONTINUE
      DO 25 I=1,NP
      F(I)=F(I)*EXP(-G(I))
   25 CONTINUE
    1 CONTINUE
      END
```

I

II

III

```
      PROGRAM TENSE(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      COMPLEX A(200),ALPHA(200),PHI(200),JH,KH
      DIMENSION R(300),C(300)
      DIMENSION X(1002),F(1002)
      DIMENSION Y(2000),FREQ(2000)
      DIMENSION CF(1000),G(1000)
      PI=4.*ATAN(1.0)
      TWOPI=2.*PI
      I=7
      L=1
      NI=93
      N=NI
      READ(5,100) (Y(I),I=1,N)
  100 FORMAT(8F10.0)
      READ(5,101) (FREQ(I),I=1,NI)
  101 FORMAT(8F10.0)
      CF(1)=FREQ(1)/110.
      FREQ(1)=Y(1)
      DO 8 I=2,NI
      CF(I)=CF(I-1)+FREQ(I)/110.
      FREQ(I)=Y(I)-Y(I-1)
    8 CONTINUE
      XMID=(CF(N)+CF(1))*0.5
      XRNG=CF(N)-CF(1)
      SCL=TWOPI/XRNG
      CALL CENTER(CF,CF,NI,XMID)
      CALL XSCALE(CF,CF,NI,SCL)
      CALL GOSPHI(CF,NI,R,C,PHI,FREQ,M,L)
      DO 15 J=2,M
      A(J)=CMPLX(R(J-1),C(J-1))
   15 CONTINUE
      A(1)=CMPLX(Y(N)-Y(1),0.)
      PRINT*,A
      K=100
      FK=K
      NF=K
      FNF=NF

      DO 1 J=2,M
      DO 13 I=1,K
      X(I)=-PI+(I-1)*TWOPI/FK
   13 CONTINUE
      CALL AUTOREG(A,J,M,K,ALPHA,PHI,JH,KH,X,F)
      PRINT*,ALPHA,KH
      X(1)=Y(1)
      F(1)=1./F(1)
      DO 14 I=2,NF
      X(I)=X(I-1)+F(I)*TWOPI/ FNF
      F(I)=1./F(I)
   14 CONTINUE

    1 CONTINUE

      END
```

I

II

III

PART II - THEORETICAL RESULTS

Before establishing the practicability of the autoregressive method, we had briefly mentioned in the Introduction two possible interpretations (namely an orthogonal polynomial interpretation and an autoregressive covariance modeling) to give some insight as to why this method could be used at all in the estimation of certain functions. We will now develop these interpretations as they really open the way to the understanding of the theoretical properties of the method.

Typically the statistical evaluation of an estimation procedure is the study of its convergence properties. This can be done quite often in two parts: first the deterministic part or study of the bias, second the stochastic part.

Paralleling Part I, we will first answer questions about the bias, that is: How good is the autoregressive method as an approximation method? Then we will consider the consistency problem so as to complete the picture of the autoregressive method as an estimation method.

In the process we will try to resolve the unanswered questions with which we concluded Part I:

- What is the best route to estimate a function?
  - Should we transform the data?
  - In the case of a density function, should we proceed directly or via the quantile function, or via the hazard function?
- How does averaging successive orders help to reduce the bias?
  - Can we explain the "odd-even" phenomenon?
- Why does symmetrizing work better in near exponential situations?

Note on our numbering system

Let a stand for a chapter,

b for a section of a chapter,

x for a subsection within a section,

y for the rank of an equation in a subsection.

All equations will be numbered (x·y) .

For reference purposes, we will use (x·y) when the reference appears within the same section (a·b) as the equation. Otherwise, we will use the complete identification (a·b·x·y).

CHAPTER 3

INTERPRETATIONS

3.1 Time series interpretation

3.1.1 Moving average process

Let $\mathbb{Z}$ be the set of all integers. Let $\{Y(t), t\in\mathbb{Z}\}$ be a complex-valued stationary time series with covariance function $R_Y(v) = E[Y(t) \cdot \overline{Y(t+v)}]$.

We say that $Y(\cdot)$ is a moving average process if there exists an orthogonal process $\{\epsilon(t), t\in\mathbb{Z}\}$ with

$$E[\epsilon(t)] = 0$$

$$R_\epsilon(v) = E[\epsilon(t) \cdot \overline{\epsilon(t+v)}] = \sigma_\epsilon^2 \delta_{v,0} \quad , \quad \sigma_\epsilon^2 > 0$$

(where $\delta_{v,0}$ is the Kronecker delta function)

such that

$$(1.1) \qquad Y(t) = \sum_{k=-\infty}^{\infty} \beta_k \epsilon(t-k) \ .$$

Define the lag operator L by

$$L^j \epsilon(t) = \epsilon(t-j)$$

then, (1.1) can be rewritten as

$$(1.2) \qquad \begin{cases} Y(t) = h(L)\,\epsilon(t) \\ \text{where } h(L) = \sum_{k=-\infty}^{\infty} \beta_k L^k \end{cases}$$

In other words we say that $Y(t)$ is the output of a filter $h(L)$ with input $\epsilon(t)$. Moreover, we have that

$$(1.3) \qquad \begin{cases} R_Y(v) = \sum_{s=-\infty}^{\infty} R_h(s)\, R_\epsilon(s - v) \\ \text{where } R_h(s) = \sum_{k=-\infty}^{\infty} \beta_k \overline{\beta_{s+k}} \quad . \end{cases}$$

Thus,

$$(1.4) \qquad R_Y(v) = \sigma_\epsilon^2 \sum_{s=-\infty}^{\infty} \beta_s \overline{\beta_{s+v}} \quad .$$

In view of the convolution formula (1.3), it is useful to consider in turn the Fourier transform of $R_Y(\cdot)$, $f_Y(\cdot)$, defined by

$$(1.5) \qquad f_Y(x) \sim \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-ivx} R_Y(v)$$

We can write it automatically as

$$(1.6) \qquad \begin{cases} f_Y(x) = f_h(x) \cdot f_\epsilon(x) \\ \text{where } f_h(x) \sim \sum_{v=-\infty}^{\infty} e^{-ivx} R_h(v) \sim \left| h(e^{ix}) \right|^2 \\ \qquad\qquad f_\epsilon(x) = \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} e^{-ivx} R_\epsilon(v) = \frac{\sigma_\epsilon^2}{2\pi} \quad . \end{cases}$$

Thus,

$$(1.7) \qquad f_Y(x) = \frac{\sigma_\varepsilon^2}{2\pi} |h(e^{ix})|^2 \quad .$$

We also have (if $F_Y(\cdot)$ is absolutely continuous)

$$(1.8) \qquad R_Y(v) = \int_{-\pi}^{\pi} e^{ivx} dF_Y(x) = \int_{-\pi}^{\pi} e^{ivx} f_Y(x)\, dx \quad .$$

Because of (1.8), we call $f_Y(\cdot)$ the spectral density of the process $Y(\cdot)$; $h(e^{ix}) \sim \sum_{k=-\infty}^{\infty} \beta_k e^{ikx}$ is called the transfer function of the filter $h(L)$, as it is the link between the time domain representation (1.2) and the frequency domain representation (1.7).

It is time that we worry a little about the meaning of all those infinite operations we have been performing.

The process $Y(\cdot)$ has finite variance if and only if $\sum_{k=-\infty}^{\infty} |\beta_k|^2 < \infty$, and then (1.1) is defined in mean square. If $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$, then

$$\sum_{v=-\infty}^{\infty} |R_Y(v)| < \left[\sum_{k=-\infty}^{\infty} |\beta_k| \right]^2 < \infty$$

and (1.5) will converge pointwise, a.e.

$$\sum_{k=-\infty}^{\infty} |\beta_k| < \infty \quad \text{implies} \quad \sum_{k=-\infty}^{\infty} |\beta_k|^2 < \infty \ ,$$

it turns out that $\sum_{k=-\infty}^{\infty} |\beta_k| < \infty$ is a sufficient condition for all our operations to be valid.

We will say that $Y(\cdot)$ has a moving average representation in terms of the past if $\beta_k = 0$ for all $k < 0$ in (1.1). We will say that $Y(\cdot)$ has a moving average representation of order q if in addition $\beta_k = 0$ for $k > q$ and $\beta_q \neq 0$ in (1.1); usually we normalize by $\beta_0 = 1$ and use the notation

$$(1.9) \qquad Y(t) = \epsilon(t) + \sum_{k=1}^{q} \beta_{kq}\, \epsilon(t - k)$$

3.1.2 Autoregressive process

Let $\{Y(t), t\epsilon\mathbb{Z}\}$ be a complex-valued stationary time series with covariance function $R_Y(\cdot)$.

We say that $Y(\cdot)$ is an autoregressive process if there exists a filter $g(L)$ and an orthogonal process $\{\eta(t), t\epsilon\mathbb{Z}\}$

$$E[\eta(t)] = 0, \quad R_\eta(v) = E[\eta(t)\,\overline{\eta(t+v)}] = \sigma_\eta^2 \delta_{v,0}, \quad \sigma_\eta^2 > 0$$

such that

$$(2.1) \qquad Y(t) = \frac{1}{g(L)}\,\eta(t)$$

where

$$g(L) = \sum_{j=0}^{\infty} \alpha_j L^j, \qquad \alpha_0 = 1 .$$

We usually write the equivalent formula

$$(2.2) \qquad Y(t) + \sum_{j=1}^{\infty} \alpha_j Y(t - j) = \eta(t)$$

Note that the filter $g(L)$ does not allow the $Y(\cdot)$ process to depend on its future, but it is conceivable that $h(L) = 1/g(L)$ would allow the future of the $\eta(\cdot)$ process to enter in (2.1). We will guard against that by asking that the roots of $g(z)$, $\{r_j^{-1}\}$, be all outside the unit circle. Then,

$$(2.3) \qquad h(L) = \frac{1}{g(L)} = \frac{1}{\prod_{j=1}^{\infty} (1 - r_j L)} = \prod_{j=1}^{\infty} \left(\sum_{k=0}^{\infty} (r_j L)^k \right)$$

which provides a moving average representation in terms of the past.

It is difficult to express $R_Y(\cdot)$ in terms of the filter $h(L)$ as in (1.4), but if we post-multiply both sides of equation (2.2) by $\overline{Y(t - k)}$, $k = 1,2,3,\ldots$, and take expectation of both sides, we obtain the following linear relations

$$(2.4) \qquad \begin{bmatrix} R_Y(0) & R_Y(1) & R_Y(2) & \cdots \\ R_Y(-1) & R_Y(0) & R_Y(1) & \cdots \\ R_Y(-2) & R_Y(-1) & R_Y(0) & \cdots \\ & \vdots & & \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \end{bmatrix} = - \begin{bmatrix} R_Y(-1) \\ R_Y(-2) \\ R_Y(-3) \\ \vdots \end{bmatrix}$$

which are called the Yule-Walker equations. When we post-multiply by $\overline{Y(t)}$ and take expectation we obtain

$$(2.5) \qquad \sum_{j=0}^{\infty} \alpha_j R_Y(j) = \sigma_\eta^2 , \qquad \alpha_0 = 1 .$$

In establishing these formulas we make use of the fact that $Y(\cdot)$ has a moving average representation in terms of the past, which implies that

$$(2.6) \qquad \begin{cases} E[\eta(t)\ \overline{Y(t-k)}] = 0 & \text{for } k > 0 , \text{ all } t \\ E[\eta(t)\ \overline{Y(t)}] = \sigma_\eta^2 \end{cases}$$

The matrix on the left of (2.4) is an infinite Toeplitz matrix.

Finally, as in the previous section, the spectral density is

$$(2.7) \qquad f_Y(x) = \frac{\sigma_\eta^2}{2\pi} \frac{1}{|g(e^{ix})|^2}$$

We say that $Y(\cdot)$ is an autoregressive process of order p if $\alpha_j = 0$ for $j > p$ and $\alpha_p \neq 0$ in (2.2). In this case we prefer to use the following notation:

$$(2.8) \qquad Y(t) + \sum_{j=1}^{p} \alpha_{jp} Y(t-j) = \eta(t)$$

The coefficients now satisfy a finite system of Yule-Walker equations

$$(2.9)\qquad \begin{cases} \begin{bmatrix} R_Y(0) & \dots & R_Y(p-1) \\ \vdots & & \vdots \\ R_Y(1-p) & \dots & R_Y(0) \end{bmatrix} \begin{bmatrix} \alpha_{1p} \\ \vdots \\ \alpha_{pp} \end{bmatrix} = - \begin{bmatrix} R_Y(-1) \\ \vdots \\ R_Y(-p) \end{bmatrix} \\ \sum_{j=0}^{p} \alpha_{jp} R_Y(j) = \sigma_\eta^2 \end{cases}$$

More precisely, the covariance function $R_Y(\cdot)$ obeys the following difference relation

$$(2.10)\qquad \sum_{j=1}^{p} \alpha_{jp} R_Y(j-v) = -R_Y(-v) \quad , \quad \text{for all} \quad v > 0 \quad .$$

It is clear that in the finite order case there is no problem in any of the operations except maybe in (2.9). But there, if $Y(\cdot)$ is stationary, the covariance function $R_Y(\cdot)$ is strictly positive definite and thus the determinant of the finite Toeplitz matrix in (2.9) is greater than zero (see Pagano (1973) for the real case). The infinite order case is treated as the limit of the finite order cases.

In the following, we shall impose that $\sum_{j=1}^{\infty} |\alpha_j|^2 < \infty$ to insure that $f_Y^{-1}(\cdot)$ is integrable as

$$Ri(0) = \frac{4\pi^2}{\sigma_\eta^2} (1 + \sum_{j=1}^{\infty} |\alpha_j|^2) = \int_{-\pi}^{\pi} \frac{dx}{f_Y(x)}$$

where we anticipate (3.2) and Table 3.1.

3.1.3 Relations between moving average and autoregressive processes

If we compare (2.7) with (1.7), we see that the spectral densities of autoregressive and moving average processes are almost inverse of each other. Let

$$(3.1)\qquad \begin{cases} f_{MA}(x) = \dfrac{\sigma_\varepsilon^2}{2\pi} \left|h(e^{ix})\right|^2 \\ f_{AR}(x) = \dfrac{\sigma_\eta^2}{2\pi} \dfrac{1}{\left|g(e^{ix})\right|^2} \end{cases}$$

Then $f_{AR}^{-1}(\cdot)$ is the spectral density of a moving average process having filter $g(L)$ and with $\sigma_\varepsilon^2 = 4\pi^2 \sigma_\eta^{-2}$. Similarly, $f_{MA}^{-1}(\cdot)$ is the spectral density of an autoregressive process having filter $h(L)$ and with $\sigma_\eta^2 = 4\pi^2 \sigma_\varepsilon^{-2}$.

To any spectral density we associate a new function $fi(\cdot) = f^{-1}(\cdot)$, the inverse spectral density (we require $f^{-1}(\cdot)$ to be integrable, so it can play the role of a spectral density), and $Ri(\cdot)$ the covinverse function (or inverse covariance function) related to $fi(\cdot)$ by

$$(3.2)\qquad Ri(v) = \int_{-\pi}^{\pi} e^{ivx} fi(x)\, dx \quad .$$

We can build the following table of relations:

Table 3.1 Relations Between Autoregressive and Moving Average Processes

Autoregressive Process		Moving Average Process	
Filter representation	$g(L)\, Y(t) = \eta(t)$	Filter representation	$Y(t) = h(L)\, \epsilon(t)$
Spectral density	$\frac{\sigma_\eta^2}{2\pi} \frac{1}{\lvert g(e^{ix}) \rvert^2}$	Inverse spectral density	$\frac{4\pi^2}{\sigma_\epsilon^2} \cdot \frac{1}{2\pi \lvert h(e^{ix}) \rvert^2}$
Inverse spectral density	$\frac{4\pi^2}{\sigma_\eta^2} \cdot \frac{\lvert g(e^{ix}) \rvert^2}{2\pi}$	Spectral density	$\sigma_\epsilon^2 \cdot \frac{\lvert h(e^{ix}) \rvert^2}{2\pi}$
Covariance function	$R(\cdot)$ obeys Yule-Walker equations	Covinverse function	$Ri(\cdot)$ obeys Yule-Walker equations
Covinverse function	$Ri(\cdot)$	Covariance function	$R(\cdot)$

3.1.4 General representation of a time series

Let $\{Y(t), t\in\mathbb{Z}\}$ be a complex-valued stationary time series with spectral density $f(\cdot)$ and covariance function $R(\cdot)$,

$$R(v) = \int_{-\pi}^{\pi} e^{ivx} f(x)\, dx \quad .$$

Then there exists an orthogonal process $\{\varepsilon(t), t\in\mathbb{Z}\}$ such that

$$(4.1) \qquad \begin{cases} Y(t) = \sum_{k=-\infty}^{\infty} \beta_k \varepsilon(t-k) \\ \text{where} \quad \sum_{k=-\infty}^{\infty} |\beta_k|^2 < \infty \quad , \end{cases}$$

and $f(\cdot)$ can be represented as

$$(4.2) \qquad \tilde{f}(x) = \sigma_\varepsilon^2 \frac{\left| \sum_{k=-\infty}^{\infty} \beta_k e^{ikx} \right|^2}{2\pi} \quad , \text{ where } f(x) < \infty \quad .$$

If $\log f(\cdot)$ is Lebesgue-integrable (which requires at least that $f(\cdot)$ be positive and finite almost everywhere on $[-\pi,\pi]$), we have that $\beta_k = 0$ for all $k < 0$, i.e. $Y(\cdot)$ has a moving average representation in terms of the past and $f(\cdot)$ is representable by

$$(4.3) \qquad \tilde{f}(x) = \sigma_\varepsilon^2 \frac{\left| \sum_{k=0}^{\infty} \beta_k e^{ikx} \right|^2}{2\pi} \quad , \text{ where } 0 \le f(x) < \infty \quad .$$

(Doob (1953), p. 577)

On the other hand, there exists a process $Y'(\cdot)$ having covariance function $Ri(\cdot)$ and spectral density $fi(\cdot) = f^{-1}(\cdot)$ with

$$Ri(v) = \int_{-\pi}^{\pi} e^{ivx} fi(x)\, dx \quad .$$

Thus there exists an orthogonal process $\{\epsilon'(t), t\in\mathbb{Z}\}$ such that

$$(4.4) \qquad \begin{cases} Y'(t) = \sum_{j=-\infty}^{\infty} \alpha_j \epsilon'(t-j) \\ \text{where} \quad \sum_{j=-\infty}^{\infty} |\alpha_j|^2 < \infty \end{cases}$$

and $fi(\cdot)$ can be represented as

$$(4.5) \qquad \tilde{fi}(x) = \sigma_\epsilon^2, \frac{\left| \sum_{j=-\infty}^{\infty} \alpha_j e^{ijx} \right|^2}{2\pi} \quad , \text{ where } fi(x) < \infty$$

As $\log fi(\cdot) = -\log f(\cdot)$, we have that $\alpha_j = 0$ for all $j < 0$ if $\log f(\cdot)$ is Lebesgue-integrable, and then $fi(\cdot)$ is representable by

$$(4.6) \qquad \tilde{fi}(x) = \sigma_\epsilon^2, \frac{\left| \sum_{j=0}^{\infty} \alpha_j e^{ijx} \right|^2}{2\pi} \quad , \text{ where } 0 \le fi(x) < \infty \quad .$$

By comparing (4.3) with (4.6), we conclude that if $\log f(\cdot)$ is Lebesgue-integrable, $f(\cdot)$ can be represented almost everywhere by either of two forms: — as a moving average spectral density

$$(4.7) \qquad \tilde{f}(x) = \sigma_\varepsilon^2 \frac{\left| \sum_{k=0}^{\infty} \beta_k e^{ikx} \right|^2}{2\pi} , \quad \sum_{k=0}^{\infty} |\beta_k|^2 < \infty$$

— as an autoregressive spectral density

$$(4.8) \qquad \tilde{f}(x) = \frac{4\pi^2}{\sigma_{\varepsilon'}^2} \frac{1}{2\pi \cdot \left| \sum_{j=0}^{\infty} \alpha_j e^{ijx} \right|^2} , \quad \sum_{j=0}^{\infty} |\alpha_j|^2 < \infty \quad .$$

3.1.5 Time series interpretation of the autoregressive method

We start with a function $R(\cdot)$ that is

(5.1) strictly positive definite

and such that

$$(5.2) \qquad R(-v) = \overline{R(v)}$$

It is well-known that such a function is a covariance function.

Then we assume that there exists a function $f(\cdot)$ defined on $[-\pi,\pi]$ such that

$$R(v) = \int_{-\pi}^{\pi} e^{ivx} f(x)\, dx$$

Furthermore, we assume the existence of a hypothetical complex-valued stationary time series $Y(\cdot)$ whose covariance function is $R(\cdot)$ and spectral density $f(\cdot)$. It is always possible to construct a Gaussian time series having zero mean and covariance function determined by $R(\cdot)$.

Assuming that

$$(5.3) \qquad -\infty < \int_{-\pi}^{\pi} \log f(x)\, dx < \infty \quad \text{and} \quad \int_{-\pi}^{\pi} \frac{dx}{f(x)} < \infty$$

we seek the autoregressive representation of $f(\cdot)$, $\tilde{f}(\cdot)$

$$(5.4) \qquad \begin{cases} \tilde{f}(x) = \dfrac{\sigma_\infty^2}{2\pi} \cdot \dfrac{1}{\left| \sum_{j=0}^{\infty} \alpha_j e^{ijx} \right|^2} , & \sum_{j=0}^{\infty} |\alpha_j|^2 < \infty \\ \text{where } \sigma_\infty^2 = \sum_{j=0}^{\infty} \alpha_j R(j) , & \text{according to (2.5)} \end{cases}$$

Now we know from subsection 3.1.2 that the α's and the function $R(\cdot)$ are related via the Yule-Walker equations (2.4). But that system is infinite; so we consider successive approximation by finite orders as in (2.9)

$$(5.5) \qquad \begin{cases} f_p(x) = \dfrac{K_p}{2\pi} \dfrac{1}{\left| 1 + \sum_{j=1}^{p} \alpha_{jp} e^{ijx} \right|^2} \\ K_p = \sum_{j=0}^{p} \alpha_{jp} R(j) , \qquad \alpha_{0p} = 1 \end{cases}$$

in the hope tnat

$$\lim_{p \to \infty} f_p(x) = \tilde{f}(x) \quad . \tag{5.6}$$

This is the question we will consider.

3.2 Orthogonal polynomial interpretation

3.2.1 Theory of orthogonal polynomials on the unit circle

Let $F(\cdot)$ be a nondecreasing bounded function with infinitely many points of increase, defined on $[-\pi,\pi]$. We denote by L_F^2 the space of measurable complex-valued functions $g(\cdot)$ such that

$$\int_{-\pi}^{\pi} |g(e^{ix})|^2 \, dF(x) < \infty \quad .$$

It is well known that L_F^2 with the following inner product

$$\Big(u(\cdot),v(\cdot)\Big)_F = \int_{-\pi}^{\pi} u(e^{ix}) \, \overline{v(e^{ix})} \, dF(x) \quad ,$$

$$u(\cdot) \text{ and } v(\cdot) \in L_F^2$$

is a Hilbert space.

If we orthogonalize in L_F^2 the set of powers $\{1,Z,Z^2,\ldots,Z^n,\ldots\}$, we obtain a set of polynomials $\{\varphi_0(\cdot),\varphi_1(\cdot),\varphi_2(\cdot),\ldots,\varphi_n(\cdot),\ldots\}$ that are uniquely determined by the conditions that

$$(1.1) \qquad \varphi_n(Z) = \sum_{j=0}^{n} a_{jn} Z^{n-j} \quad , \quad a_{0n} > 0 \, , \text{ for all } n$$

$$(1.2) \qquad \Big(\varphi_j(\cdot),\varphi_k(\cdot)\Big)_F = \delta_{jk} \, , \quad \text{for all } j \text{ and } k \quad .$$

In order to construct the polynomial $\varphi_n(\cdot)$, we define the characteristic sequence $R(\cdot)$ by

$$R(v) = \int_{-\pi}^{\pi} e^{ivx}\, dF(x) \quad , \qquad v = 0, \pm 1, \pm 2, \ldots$$

Note that $R(-v) = \overline{R(v)}$. The normal equations (1.2) can be replaced for $\varphi_n(\bullet)$ by

$$(1.3) \qquad \begin{cases} \int_{-\pi}^{\pi} \varphi_n(e^{ix})\, e^{-ijx}\, dF(x) = 0 \quad , \qquad j = 0,1,\ldots,n-1 \\ \left(\varphi_n(\bullet),\varphi_n(\bullet)\right)_F = 1 \end{cases}$$

which we rewrite as

$$(1.4) \qquad \begin{cases} \sum_{\ell=0}^{n} a_{\ell n} R(n-\ell-j) = 0 \quad , \qquad j = 0,1,\ldots,n-1 \\ a_{0n} \sum_{\ell=0}^{n} a_{\ell n} R(-\ell) = 1 \end{cases}$$

In matrix form this system is equivalent to

$$(1.5) \qquad \begin{bmatrix} R(0) & R(1) & \ldots & R(n) \\ \vdots & & & \vdots \\ R(-n+1) & R(-n+2) & \ldots & R(1) \\ R(-n) & R(-n+1) & \ldots & R(0) \end{bmatrix} \begin{bmatrix} a_{0n} \bullet a_{nn} \\ \vdots \\ a_{0n} \bullet a_{1n} \\ a_{0n} \bullet a_{0n} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

In view of the multiplicative effect a_{0n} has on the coefficients in (1.5), we can reduce the system by setting $a_{0n} = 1$ and then recover the true value of a_{0n} via the normalization $\left(\varphi_n(\bullet),\varphi_n(\bullet)\right)_F = 1$, as follows:

$$
(1.6)\qquad
\begin{cases}
\begin{bmatrix} R(0) & R(1) & \dots & R(n-1) \\ \vdots & & & \vdots \\ R(-n+2) & R(-n+3) & \dots & R(1) \\ R(-n+1) & R(-n+2) & \dots & R(0) \end{bmatrix}
\begin{bmatrix} a^*_{nn} \\ \vdots \\ a^*_{2n} \\ a^*_{1n} \end{bmatrix}
= -\begin{bmatrix} R(n) \\ \vdots \\ R(2) \\ R(1) \end{bmatrix} \\[2ex]
a^2_{0n} = \dfrac{1}{\int_{-\pi}^{\pi} \left| e^{inx} + \sum_{j=1}^{n} a^*_{jn} e^{i(n-j)x} \right|^2 dF(x)}
\end{cases}
$$

Thus $\varphi_n(Z) = a_{0n}(Z^n + a^*_{1n} Z^{n-1} + \dots + a^*_{nn})$, $a_{0n} > 0$.

Consider now the subspace, of L^2_F , generated by $\left(\varphi_0(\cdot),\varphi_1(\cdot),\dots,\varphi_n(\cdot)\right)$ and denoted by L^2_n .

L^2_n is a reproducing kernel Hilbert space, that is there exists a function $K_n(\cdot,\cdot)$ of two complex variables such that

$$
(1.7)\qquad
\begin{aligned}
&K_n(\cdot,y) \in L^2_n \\
&K_n(\cdot,y) = \sum_{j=0}^{n} k_{jn}(y)\, \varphi_j(\cdot)
\end{aligned}
$$

$$
(1.8)\qquad \left(g(\cdot),K_n(\cdot,y)\right)_F = g(y) \quad , \quad \text{for all } g(\cdot) \in L^2_n \quad .
$$

We can obtain an explicit representation for $K_n(\cdot,y)$ by solving the following normal equations:

$$(1.9) \qquad \left(\varphi_j(\cdot), K_n(\cdot,y)\right)_F = \overline{\varphi_j(y)} \quad , \qquad j = 0,1,\ldots,n$$

to yield that

$$(1.10) \qquad K_n(\cdot,y) = \sum_{j=0}^{n} \overline{\varphi_j(y)}\, \varphi_j(\cdot) \quad .$$

At the same time, we have verified the reproducing property (1.8) because any $g(\cdot) \in L_n^2$ can be written as

$$g(\cdot) = \sum_{j=0}^{n} g_j \varphi_j(\cdot)$$

It is clear also that $K_n(\cdot,y)$ is a polynomial of degree n (for fixed y). So we may seek its equivalent polynomial representation

$$(1.11) \qquad K_n(Z,y) = \sum_{j=0}^{n} a_{jn}(y)\, Z^j \quad .$$

Let $h_j(Z) = Z^j$, $j = 0,1,\ldots,n$. Then, by (1.8),

$$\left(h_j(\cdot), K_n(\cdot,y)\right)_F = h_j(y) = y^j \quad , \qquad j = 0,1,\ldots,n$$

This can be rewritten as

$$(1.12)\qquad \begin{cases} \sum_{\ell=0}^{n} \overline{a_{\ell n}(y)}\, R(j-\ell) = y^j, & j = 1,\dots,n \\ \sum_{\ell=0}^{n} \overline{a_{\ell n}(y)}\, R(-\ell) = 1 . \end{cases}$$

In particular if we set $y = 0$, we get the following system

$$(1.13)\qquad \begin{cases} \sum_{\ell=0}^{n} \overline{a_{\ell n}(0)}\, R(j-\ell) = 0, & j = 1,\dots,n \\ \sum_{\ell=0}^{n} \overline{a_{\ell n}(0)}\, R(-\ell) = 1 . \end{cases}$$

This system is equivalent to (1.4) with the following identification:

(1.13)		(1.4)
j	$\longleftrightarrow$	$n-j$
$\overline{a_{\ell n}(0)}$	$\longleftrightarrow$	$a_{0n} \cdot a_{\ell n}$

Thus while $\varphi_n(Z) = \sum_{j=0}^{n} a_{jn} Z^{n-j}$, we have that

$$(1.14)\qquad K_n(Z,0) = a_{0n} \sum_{j=0}^{n} \overline{a_{jn}}\, Z^j ,$$

$$(1.15)\qquad K_n(0,0) = a_{0n}^2 = \sum_{j=0}^{n} |\varphi_j(0)|^2 = \sum_{j=0}^{n} |a_{jj}|^2 .$$

Let $\varphi_n^*(Z) = K_n(Z,0)/a_{0n}$, then

$$(1.16)\qquad (\varphi_n^*(\cdot),\varphi_n^*(\cdot))_F = \frac{K_n(0,0)}{a_{0n}^2} = 1 .$$

3.2.2 Some interesting properties

A - Extremal properties

The polynomials $\varphi_n(Z)$ and $K_n(Z,0)$ are also related through a minimum norm approximation problem. Suppose we want to find the best approximation to Z^n in the space L^2_{n-1}, $\sum_{j=0}^{n-1} c_j^* Z^j$, such that

$$\|Z^n - \sum_{j=0}^{n-1} c_j^* Z^j\|_F^2 = \int_{-\pi}^{\pi} |e^{inx} - \sum_{j=0}^{n-1} c_j^* e^{ijx}|^2 dF(x) \quad \text{is a minimum.}$$

We know that our answer will be the projection of Z^n on L^2_{n-1}.

Let $g^*(Z) = Z^n - \sum_{j=0}^{n-1} c_j^* Z^j$. We certainly have that $g^*(Z) = \sum_{j=0}^{n} \beta_j \varphi_j(Z)$, and moreover we know that $\beta_n = \frac{1}{a_{0n}}$, for the coefficient of Z^n is equal to 1. It follows that

$$\|g^*(\cdot)\|_F^2 = \sum_{j=0}^{n} |\beta_j|^2 \geq |\beta_n|^2 = \frac{1}{a_{0n}^2} .$$

But $\frac{\varphi_n(Z)}{a_{0n}}$ attains this lower bound. Thus, $\frac{\varphi_n(Z)}{a_{0n}}$ satisfies all the requirements for $g^*(Z)$ and is our answer. So

$$c^*_{n-j} = -\frac{a_{jn}}{a_{0n}} .$$

If $\int_{-\pi}^{\pi} |e^{inx} - \sum_{j=0}^{n-1} c_j^* e^{ijx}|^2 dF(x)$ is a minimum, the same can be said of the equivalent form

$$\int_{-\pi}^{\pi} |e^{inx}|^2 \, |e^{-inx} - \sum_{j=0}^{n-1} \overline{c_j^*} \, e^{-ijx}|^2 \, dF(x)$$

or

$$\int_{-\pi}^{\pi} \left| 1 - \sum_{j=0}^{n-1} \overline{c_j^*} e^{i(n-j)x} \right|^2 dF(x)$$

Now,

$$1 - \sum_{j=0}^{n-1} \overline{c_j^*} z^{n-j} = 1 + \sum_{j=1}^{n} \frac{\overline{a_{jn}}}{a_{0n}} z^j = \frac{K_n(Z,0)}{K_n(0,0)} \quad .$$

In other words, $z^n - \frac{\varphi_n(z)}{a_{0n}}$ is the projection of z^n on L^2_{n-1} and $1 - \frac{K_n(Z,0)}{K_n(0,0)}$ is the projection of 1 on the subspace of L^2_n generated by $\{Z,\ldots,z^n\}$; $\frac{1}{K_n(0,0)}$ is the squared distance between z^n and its projection on L^2_{n-1} .

B - Recurrence relations

Let H be a Hilbert space with inner product denoted by $(f,g)_H$, for f,g in H . We denote by $f(\cdot|g)$ the projection of f on the subspace of H generated by g . By analogy with regression theory, we extend the notion of partial correlation coefficient to the context of a general Hilbert space: for any elements $f,g,f_1,\ldots,f_n$ in H , the generalized partial correlation coefficient between f and g , given $f_1,\ldots,f_n$ is defined as

$$(2.1) \quad r(f,g|f_1,\ldots,f_n) = \frac{\left(f - f(\cdot|f_1,\ldots,f_n), g - g(\cdot|f_1,\ldots,f_n)\right)_H}{\|f - f(\cdot|f_1,\ldots,f_n)\|_H \cdot \|g - g(\cdot|f_1,\ldots,f_n)\|_H}$$

where $\|h\|_H = (h,h)_H^{\frac{1}{2}}$ is the norm of h in H .

Then, for any f, g and h in H, we have that

$$(2.2) \qquad f(\cdot|g,h) = f(\cdot|h) + f(\cdot|g - g(\cdot|h)) \quad .$$

To prove this, it is sufficient to note that h and $g - g(\cdot|h)$ generate the same subspace as g and h, as they form a basis for that subspace.

We make two applications of the identity (2.2) to the case of our Hilbert space L_F^2. We represent z^n by f_n, for $n = 0,1,2,\ldots$

The first application is

$$(2.3) \qquad f_n(\cdot|f_0,\ldots,f_{n-1}) = f_n(\cdot|f_1,\ldots,f_{n-1}) + f_n(\cdot|f_0 - f_0(\ f\ ,\ldots,f_{n-1}))$$

which translates into

$$(2.4) \qquad z^n - \frac{\varphi_n(z)}{a_{0n}} = z^n - \frac{z\varphi_{n-1}(z)}{a_{0,n-1}} + \alpha \frac{K_{n-1}(z,0)}{K_{n-1}(0,0)} \quad , \quad n = 1,2,\ldots$$

wnere $\alpha = \dfrac{-a_{nn}}{a_{0n}}$ is found by equating the constant term on both sides of (2.4).

The second is

$$(2.5) \qquad f_0(\cdot|f_1,\ldots,f_n) = f_0(\cdot|f_1,\ldots,f_{n-1}) + f_0(\cdot|f_n - f_n(\cdot|f_1,\ldots,f_{n-1}))$$

which translates into

$$(2.6)\qquad 1 - \frac{K_n(Z,0)}{K_n(0,0)} = 1 - \frac{K_{n-1}(Z,0)}{K_{n-1}(0,0)} + \beta \frac{Z\varphi_{n-1}(Z)}{a_{0,n-1}} \quad , \; n = 1,2,\ldots$$

where $\beta = -\dfrac{\overline{a_{nn}}}{a_{0n}}$ is found by equating the coefficient of Z^n on both sides of (2.6).

Thus,

$$(2.7)\qquad \frac{\varphi_n(Z)}{a_{0n}} = \frac{Z\varphi_{n-1}(Z)}{a_{0,n-1}} + \frac{a_{nn}}{a_{0n}} \frac{K_{n-1}(Z,0)}{K_{n-1}(0,0)}$$

and

$$(2.8)\qquad \frac{K_n(Z,0)}{K_n(0,0)} = \frac{K_{n-1}(Z,0)}{K_{n-1}(0,0)} + \frac{\overline{a_{nn}}}{a_{0n}} \frac{Z\omega_{n-1}(Z)}{a_{0,n-1}} \quad .$$

Moreover, we can give the following interpretation to the coefficient $\dfrac{\overline{a_{nn}}}{a_{0n}}$:

$$(2.9)\qquad \frac{\overline{a_{nn}}}{a_{Cn}} = \begin{cases} -(f_0,f_1)_H/(f_1,f_1)_H \quad , & n = 1 \\ -r(f_0,f_n|f_1,\ldots,f_{n-1}) \; , & n = 2,3,\ldots \end{cases} .$$

We use the relation (2.5) in the modified form

$$(2.10)\quad f_0(\cdot|f_1,\ldots,f_n) = f_0(\cdot|f_1,\ldots,f_{n-1}) + \beta\left[f_n - f_n(\cdot|f_1,\ldots,f_{n-1})\right]$$

together with the basic property of the projection

$$(2.11)\qquad (f_0,f_n)_H = (f_0(\cdot|f_1,\ldots,f_n),\, f_n)_H$$

to obtain that

$$(2.12) \qquad \beta = \frac{(f_0 - f_0(\cdot|f_1,\ldots,f_{n-1}),\ f_n)_H}{(f_n - f_n(\cdot|f_1,\ldots,f_{n-1}),\ f_n)_H} \quad .$$

The second element f_n in each inner product can be replaced by $f_n - f_n(\cdot|f_1,\ldots,f_{n-1})$ and finally it can be verified directly that

$$\|f_0 - f_0(\cdot|f_1,\ldots,f_{n-1})\|_H = \|f_n - f_n(\cdot|f_1,\ldots,f_{n-1})\|_H \ ,$$

thus completing the proof of (2.9).

In appendix 3.A.1, we use (2.8) and (2.12) to obtain a recursive algorithm for the computation of $K_n(Z,0)$.

C - Asymptotic properties

Theorem 3.1:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{K_n(0,0)}{|K_n(e^{ix},0)|^2} e^{ikx}\,dx = \int_{-\pi}^{\pi} e^{ikx}\,dF(x) \ , \quad k = 0, \pm 1, \ldots, \pm n$$

Proof: See Geronimus (1961), p. 12.

In what sense is this result useful to us?

We know from real analysis that any monotone function $F(\cdot)$ can be written as a sum

$$F(\cdot) = F_{ac}(\cdot) + F_s(\cdot)$$

where — $F_{ac}(\cdot)$ is absolutely continuous with respect to Lebesgue measure

— $F_s(\cdot)$ is the sum of a step-function with a singular function

Let $f(\cdot)$ be the derivative of the absolutely continuous part of $F(\cdot)$.

In general, Theorem 1 means that $F(x)$ and $\int_{-\pi}^{x} \frac{K_p(0,0)}{|K_p(e^{i\theta},0)|^2}\, d\theta$ have the same first $(p+1)$ elements in their characteristic sequence. In the case $F(\cdot)$ is absolutely continuous, the Fourier series of the difference

$$f(x) - \frac{1}{2\pi} \frac{K_p(0,0)}{|K_p(e^{ix},0)|^2}$$

is of the form

$$\sum_{|k|>p} b_k e^{-ikx} \quad .$$

Now, what happens in the limit as $p \to \infty$? We have the following theorems that we take again from Geronimus (1961), except that the notation is changed.

Theorem 3.2:

$$0 < \lim_{p\to\infty} K_p(0,0) = \sum_{j=0}^{\infty} |\varphi_j(0)|^2 = K(0,0) < \infty$$

if and only if $\log f(\cdot)$ is Lebesgue-integrable, that is if and only if $\int_{-\pi}^{\pi} \log f(x)\, dx > -\infty$.

Proof: See Geronimus (1961), p. 14-17.

Theorem 3.3:

$$\pi(Z) = \lim_{p\to\infty} \frac{K_p(Z,0)}{\sqrt{K_p(0,0)}} = \lim_{p\to\infty} \varphi_p^*(Z) = \frac{1}{\sqrt{K(0,0)}} \sum_{j=0}^{\infty} \overline{\varphi_j(0)}\, \varphi_j(Z) \ , \quad |Z| < 1$$

if and only if $\log f(\bullet)$ is Lebesgue-integrable. The convergence is uniform in any closed region $|Z| \le r < 1$.

Proof: See Geronimus (1961), Chapter II.

Theorem 3.4:

$$\frac{1}{\pi(e^{ix})} = \lim_{r\to 1^-} \frac{1}{\pi(re^{ix})} \qquad \text{exists almost everywhere in } [-\pi,\pi] \ .$$

Also,

$$f(x) = \frac{1}{2\pi} \frac{1}{|\pi(e^{ix})|^2} \ , \quad \text{a.e., in } [-\pi,\pi] \ .$$

Finally, let $E = \{x\in[-\pi,\pi] \ , \ 0 < f(x) < \infty\}$ and define

$$\pi_E(x) = \begin{cases} \pi(e^{ix}) & , \quad x \in E \\ 0 & , \quad x \notin E \end{cases} .$$

Then, $\pi_E(x)$ has the following expansion in terms of the orthogonal

polynomials $\{\varphi_n(\cdot)\}_{n=0}^{\infty}$,

$$\pi_E(x) \sim \frac{1}{\sqrt{K(0,0)}} \sum_{j=0}^{\infty} \overline{\varphi_j(0)}\ \varphi_j(e^{ix}) \quad ,$$

which converges in L_F^2 .

Proof: See Geronimus (1961), Chapter II.

Theorem 3.5:

Let

$$\delta_p = \left\| \pi_E(\cdot) - \frac{\sqrt{K_p(0,0)}}{\sqrt{K(0,0)}}\ \varphi_p^*(\cdot) \right\|_F \quad ,$$

then,

$$\delta_p = \frac{1}{\sqrt{K(0,0)}} \sum_{j=p+1}^{\infty} |\varphi_j(0)|^2$$

and $\lim_{p \to \infty} \delta_p = 0$.

Proof: The expansion of $\pi_E(\cdot)$ in terms of $\{\varphi_n(\cdot)\}_{n=0}^{\infty}$ is, from Theorem 3.4.

$$\pi_E(x) \sim \frac{1}{\sqrt{K(0,0)}} \sum_{j=0}^{\infty} \overline{\varphi_j(0)}\ \varphi_j(e^{ix}) \quad .$$

The expansion of $\varphi_p^*(\cdot)$ is

$$\varphi_p^*(e^{ix}) = \frac{1}{a_{0p}} \cdot \sum_{j=0}^{p} \overline{\varphi_j(0)}\, \varphi_j(e^{ix})$$

and

$$\frac{\sqrt{K_p(0,0)}}{\sqrt{K(0,0)}}\ \varphi_p^*(e^{ix}) = \frac{1}{\sqrt{K(0,0)}} \sum_{j=0}^{p} \overline{\varphi_j(0)}\, \varphi_j(e^{ix})$$

as

$$a_{0p} = \sqrt{K_p(0,0)} \quad .$$

Now by Plancherel's theorem

$$\|\pi_E(\cdot) - \frac{\sqrt{K_p(0,0)}}{\sqrt{K(0,0)}}\ \varphi_p^*(e^{ix})\|_F = \|\frac{1}{\sqrt{K(0,0)}}\left(\sum_{j=0}^{\infty} \overline{\varphi_j(0)}\, \varphi_j(e^{ix}) - \sum_{j=0}^{p} \overline{\varphi_j(0)}\, \varphi_j(e^{ix})\right)\|_F$$

So,

$$\delta_p = \frac{1}{\sqrt{K(0,0)}} \sum_{j=p+1}^{\infty} |\varphi_j(0)|^2$$

And, by Theorem 3.2

$$\lim_{p\to\infty} \delta_p = 0 \quad .$$

Theorem 3.6:

If $F(\cdot)$ is absolutely continuous in $[-\pi,\pi]$, $0 < m \le f(x) \le M$, a.e., in $[-\pi,\pi]$ and $f(\cdot) \in \mathrm{Lip}(1/2,2)$, then

$$|\varphi_n^*(z)| \leq C \qquad , \qquad |z| \leq 1$$

Proof: See Geronimus (1961), Chapter III, Theorem 3.8.

Note: $f(\cdot) \in \text{Lip}(\alpha,p)$ if

$$w_p(\delta;f) \underset{|h|<\delta}{\sup} \left(\int_{-\pi}^{\pi} |f(x+h) - f(x)|^p \, dx\right)^{1/p} = O(\delta^{\alpha}) \quad .$$

We get the same result as Theorem 3.6 if

$$R(v) = \int_{-\pi}^{\pi} e^{ivx} f(x) \, dx = O(v^{-1})$$

Theorem 3.7:

If $\int_{-\pi}^{\pi} \log F'(x) \, dx > -\infty$, $\delta_p = o\left(\frac{1}{p}\right)$

and $F(x_2) - F(x_1) \geq m(x_2 - x_1)$, for $m > 0$ and $-\pi \leq x_1 < x_2 \leq \pi$,

then $\lim_{p\to\infty} \varphi_p^*(e^{ix}) = \pi(e^{ix})$, a.e., in $[-\pi,\pi]$.

Proof: See Geronimus (1961), Chapter V, Theorem 5.1 .

In the case $F(\cdot)$ is absolutely continuous and $f(\cdot)$ is bounded above and below, it is sufficient to have $w_2(\delta;f) = o(\sqrt{\delta})$. If $w_2(\delta;f) = O(\delta^{\alpha})$, $\alpha > \frac{1}{2}$, then the convergence is uniform (Geronimus (1961), Theorem 5.2).

We can reformulate the assumptions of some of these theorems in terms of the sequence of partial correlation coefficients $\{\alpha_{nn}\}$ by noting the following results:

Theorem 3.8:

The condition $|\alpha_{nn}| < 1$, for $n = 1,2,\ldots,$ determines the entire set of orthogonal polynomials $\{\varphi_n(\cdot)\}_{n=0}^{\infty}$ up to a multiplicative constant $(\varphi_0(\cdot) = 1)$ and thus determines a function $F(\cdot)$, bounded nondecreasing with infinitely many points of increase.

Proof: See Geronimus (1961), Chapter VIII, Theorem 8.1.

Theorem 3.9:

$\log f(\cdot)$ is Lebesgue-integrable if and only if $\sum_{n=1}^{\infty} |\alpha_{nn}|^2 < \infty$.

Proof: See Geronimus (1961), Chapter VIII, Theorem 8.2.

Theorem 3.10:

If $|\alpha_{nn}| < \frac{3}{n \log n}$, for n large enough, we have that at all x where $f(x) > 0$

$$\lim_{p \to \infty} \varphi_p^*(e^{ix}) = \pi(e^{ix}) .$$

Proof: See Geronimus (1961), Chapter VIII, Theorem 8.4.

Theorem 3.11:

If $|\alpha_{nn}| < 1$ for all n, then

$$|\varphi_p(z)| \leq a_{0p} \exp\left(\sum_{j=1}^{p} |\alpha_{jj}|\right) , \qquad |z| \leq 1$$

Proof: See Geronimus (1961), Chapter VIII, Theorem 8.3.

There does not seem to be any condition on the $\{\alpha_{nn}\}$ that would yield the equivalent of Theorem 3.6. Indeed, we have the following:

Theorem 3.12:

If $\sum_{n=1}^{\infty} |\alpha_{nn}| < \infty$, then

$|\varphi_p(z)| \leq C$, for $|z| \leq 1$, and all p

$F(\cdot)$ is absolutely continuous in $[-\pi,\pi]$,

$f(x) \geq m > 0$,

$f(\cdot)$ is continuous and

$$|\varphi_p^*(z) - \pi(z)| \leq C' \cdot \sum_{j=p}^{\infty} |\alpha_{jj}| , \qquad |z| \leq 1$$

Proof: The first assertion is a direct consequence of Theorem 3.11. For the other assertions, see Geronimus (1961), Chapter VIII, Theorem 8.5.

So, at the same time as the boundedness of $\{\omega_n(\cdot)\}_{n=0}^{\infty}$ and $\{\varphi_n^*(\cdot)_{n=0}^{\infty}$, we get the convergence of $\varphi_p^*(e^{ix})$ to $\pi(e^{ix})$. This is equivalent to the combination of Theorems 3.6 and 3.7 with uniform convergence.

3.2.3 Orthogonal polynomial interpretation of the autoregressive method

If we want to approximate the derivative $f(\cdot)$ of a bounded nondecreasing function $F(\cdot)$, we form the characteristic sequence

$$R(v) = \int_{-\pi}^{\pi} e^{ivx}\, dF(x), \qquad v = 0,1,\ldots,n$$

from which we obtain the orthogonal polynomials $\{\varphi_0(\cdot),\varphi_1(\cdot),\ldots,\varphi_n(\cdot)\}$ and the related kernel functions $\{K_0(\cdot,0),K_1(\cdot,0),\ldots,K_n(\cdot,0)\}$.

Under the assumption that $\log f(\cdot)$ is Lebesgue-integrable, we have that

$$f_p(x) = \frac{1}{2\pi} \frac{K_p(0,0)}{|K_p(e^{ix},0)|^2}$$

is an approximator of $f(x)$.

In the next chapter, we give more precision to that affirmation.

The estimation problem would be similar save for the estimation of the $R(\cdot)$ sequence usually through the use of a crude estimator of $F(\cdot)$.

3.3 Correspondences between the two interpretations

The autoregressive approximator of order p

$$(0.1) \qquad f_p^{(1)}(x) = \frac{K_p}{2\pi} \cdot \frac{1}{\left|1 + \sum_{j=1}^{p} \alpha_{jp} e^{ijx}\right|^2}$$

and the approximator of order p we obtain from the orthogonal polynomial theory

$$(0.2) \qquad f_p^{(2)}(x) = \frac{K_p(0,0)}{2\pi} \cdot \frac{1}{|K_p(e^{ix},0)|^2} = \frac{K_p(0,0)}{2\pi} \cdot \frac{1}{\left|a_{0p} \sum_{j=0}^{p} \overline{a_{jp}} e^{ijx}\right|^2}$$

are equal.

Indeed in the time series case, the parameters $\{\alpha_{jp}\}_{j=1}^{p}$ and K_p are related by the following equations

$$(0.3) \qquad \begin{cases} \sum_{j=1}^{p} \alpha_{jp} R(j-\ell) = -R(-\ell), & \ell = 1,\ldots,p \\ \sum_{j=0}^{p} \alpha_{jp} R(j) = K_p \end{cases}$$

(see equation 3.1.2.9)

In the orthogonal polynomial case, we can rewrite the system (3.2.1.13) as

$$(0.4)\qquad \begin{cases} \sum_{j=1}^{p} \overline{a_{jp}}\, R(j-\ell) = -a_{0p}\, R(-\ell) \\ a_{0p} \sum_{j=0}^{p} \overline{a_{jp}}\, R(j) = 1 \end{cases}$$

The following identifications provide the equivalence of the two systems:

$$(0.5)\qquad \begin{cases} \alpha_{0p} = 1 \\ \alpha_{jp} = \dfrac{\overline{a_{jp}}}{a_{0p}} \\ K_p = \dfrac{1}{a_{0p}^2} = \dfrac{1}{K_p(0,0)} \end{cases}$$

It follows that all the properties that were established in subsection 3.2.2 have a time series interpretation. We note first that from the definition of $R(\cdot)$ in both interpretations, we can make the identification of $Y(t-v)$ with Z^v. Then, from subsection 3.2.2 part A, the best (minimum mean-square error) linear predictor of $Y(t-p)$ given $Y(t-p+1),\ldots,Y(t)$, denoted by $Y(t-p|t-p+1,\ldots,t)$ is

$$(0.6)\qquad Y(t-p|t-p\quad 1,\ldots,t) = -\sum_{j=1}^{p} \overline{\alpha_{jp}}\, Y(t-p+j)\ ,$$

and the best linear predictor of $Y(t)$ given $Y(t-1),\ldots Y(t-p)$ is

$$(0.7) \qquad Y(t|t-1,\ldots,t-p) = -\sum_{j=1}^{p} \alpha_{jp}\, Y(t-j) \quad .$$

The recurrence relations of subsection 3.2.2 part B become

$$(0.8) \qquad \begin{aligned} Y(t-p|t-p+1,\ldots,t) &= Y(t-p|t-p+1,\ldots,t-1) \\ &\quad - \overline{\alpha}_{pp} \sum_{j=0}^{p-1} \alpha_{j,p-1}\, Y(t-j) \end{aligned}$$

and

$$(0.9) \qquad \begin{aligned} Y(t|t-1,\ldots,t-p) &= Y(t|t-1,\ldots,t-p+1) \\ &\quad - \alpha_{pp} \sum_{j=0}^{p-1} \overline{\alpha}_{j,p-1}\, Y(t-p+j) \quad . \end{aligned}$$

$(-\alpha_{pp})$ is the partial correlation coefficient between $Y(t)$ and $Y(t-p)$ given $Y(t-1,\ldots,Y(t-p+1)$, and by evaluating (3.2.2.12), we obtain that

$$(0.10) \qquad -\alpha_{pp} = \frac{\sum_{j=0}^{p-1} \alpha_{j,p-1}\, R(j-p)}{\sum_{j=0}^{p-1} \alpha_{j,p-1}\, R(j)} = \frac{\sum_{j=0}^{p-1} \alpha_{j,p-1}\, R(j-p)}{K_{p-1}} \quad .$$

K_p is the mean-square prediction error when we use $Y(t|t-1\ldots,t-p)$ to predict $Y(t)$.

Finally, the theorems of subsection 3.2.2 part C apply directly.

In particular, Theorem 3.1 can be rephrased to say

$$(0.11) \qquad \int_{-\pi}^{\pi} f_p^{(j)}(x)\, e^{ikx}\, dx = \int_{-\pi}^{\pi} e^{ikx}\, dF(x) \ , \quad |k| \le p \ , \quad j = 1,2$$

and Theorem 3.2 implies that

$$(0.12) \qquad K_p \text{ decreases to } K = \frac{1}{K(0,0)} \text{ as } p \to \infty \ .$$

Appendix

3.A.1 Recursive algorithm

We provide a recursive algorithm to compute

$$f_p(x) = \frac{K_p}{2\pi} \cdot \frac{1}{\left|1 + \sum_{j=1}^{p} \alpha_{jp} e^{ijx}\right|^2}$$

given K_{p-1}, $\{\alpha_{j,p-1}\}_{j=1}^{p-1}$ and the sequence $R(\cdot)$.

From (3.3.0.10), which is the equivalent of (3.2.2.12), we have that

$$\alpha_{pp} = -\sum_{j=0}^{p-1} \alpha_{j,p-1} R(j-p)/K_{p-1} \quad . \tag{1.1}$$

From (3.3.0.9), which is the equivalent of (3.2.2.8), we have that

$$\alpha_{jp} = \alpha_{j,p-1} + \alpha_{pp} \overline{\alpha_{p-j,p-1}} \quad , \qquad j = 1,\ldots,p-1 \quad . \tag{1.2}$$

Finally, from (3.1.5.5), we have that

$$K_p = \sum_{j=0}^{p} \alpha_{jp} R(j) \quad , \qquad \alpha_{0p} = 1 \quad . \tag{1.3}$$

Using (1.2), we obtain that

$$K_p = K_{p-1} + \alpha_{pp} \sum_{j=1}^{p} \overline{\alpha_{p-j,p-1}} R(j)$$

$$= K_{p-1} - \frac{\left| \sum_{j=1}^{p} \alpha_{j,p-1} R(j-p) \right|^2}{K_{p-1}}$$

The initial conditions are simply

$$(1.4) \qquad K_0 = R(0) \quad , \qquad \alpha_{00} = 1 \quad .$$

We have written a FORTRAN subroutine that computes $f_p(\cdot)$. This subroutine is called AUTOREG and can be found on the next page.

```
      SUBROUTINE AUTOREG(A,K,M,NP,ALPHA,PHI,JH,KH,X,F)
C
C     COMPUTES THE COEFFICIENTS ALPHA(.), ACCORDING TO
C     A RECURSIVE ALGORITHM, AND THE CORRESPONDING
C     FUNCTION F(.) AT THE POINTS X(.)
C     INPUT-
C          A = VECTOR OF COMPLEX FOURIER TRANSFORM,
C              OF DIMENSION AT LEAST M
C          K : (K-1) IS THE ACTUAL ORDER OF THE SCHEME
C              BEING COMPUTED, K.GE.2
C          M : (M-1) IS THE MAXIMUM ORDER OF SCHEME
C              TO BE COMPUTED
C          X = VECTOR OF VALUES AT WHICH F(.) IS TO
C              BE COMPUTED, OF DIMENSION NP; -PI.LE.X..E.PI
C     OUTPUT-
C          ALPHA = VECTOR OF COEFFICIENTS DEFINING THE
C              APPROXIMATING FUNCTION,HAS TO BE DIMEN-
C              SIONED AT LEAST M
C          F = VECTOR OF VALUES OF THE APPROXIMATING
C              FUNCTION, OF DIMENSION NP
C     ALPHA, PHI, JH, KH ARE USED RECURSIVELY, THAT IS
C     THEIR VALUES AT OUTPUT FOR K ARE USED AS INPUT
C     FOR (K+1)
C
      DIMENSION X(1),F(1)
      COMPLEX A(1),ALPHA(1),PHI(1),JH,KH,G
C
      TWOPI=6.2831853
      JH=CMPLX(0.,0.)
      ALPHA(1)=CMPLX(1.,0.)
      L=K-1
      PHI(L)=CMPLX(1.,0.)
      IF(K.EQ.2) KH=CONJG(A(1))
      DO 4 I=1,L
    4 JH=JH+CONJG(A(I+1))*PHI(I)
      G=-JH/KH
      ALPHA(K)=G
      IF(L.EQ.1) GO TO 5
      DO 2 I=2,L
    2 ALPHA(I)=ALPHA(I)+G*PHI(I-1)
    5 CONTINUE
      DO 3 I=1,K
    3 PHI(I)=CONJG(ALPHA(K+1-I))
      KH=KH-JH*CONJG(JH)/CONJG(KH)
      CC=KH/TWOPI
      DO 11 I=1,NP
      G=CMPLX(1.,0.)
      DO 12 J=2,K
      FJ=J-1
   12 G=G+CEXP(CMPLX(0.,X(I)*FJ))*ALPHA(J)
      F(I)=CC/(G*CONJG(G))
   11 CONTINUE
      RETURN
      END
```

CHAPTER 4

BIAS STUDY

The Autoregressive Mechod as an Approximation Method

The properties of the autoregressive approximator $f_p(\cdot)$ depend in a large measure on the function $f(\cdot)$ being approximated. In this chapter, we study the effect various hypotheses concerning $f(\cdot)$ have on the behavior of the bias function

$$b_p(\cdot) = f(\cdot) - f_p(\cdot)$$

especially the rate at which it goes to zero.

In this chapter, we follow Geronimus (1961) closely, except in Section 4.2.

4.1 Autoregressive representation

4.1.1 Convergence "in the mean"

We start our study with the case where $F(\cdot)$ is absolutely continuous and where its derivative $f(\cdot)$ has the following properties

$$(1.1) \qquad \begin{cases} 0 < \int_{-\pi}^{\pi} f(x)\,dx = R(0) < \infty \\ 0 < \int_{-\pi}^{\pi} \frac{1}{f(x)}\,dx = Ri(0) < \infty. \\ -\infty < \int_{-\pi}^{\pi} \log f(x)\,dx < \log R(0) < \infty \end{cases}$$

Under these conditions, $f(\cdot)$ has the following autoregressive representation

$$f(x) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{\left|1 + \sum_{j=1}^{\infty} \alpha_j e^{ijx}\right|^2}, \qquad \begin{matrix} \sigma^2 > 0 \\ \sum_{j=1}^{\infty} |\alpha_j|^2 < \infty \end{matrix}$$

and

$$f_p(x) = \frac{K_p}{2\pi} \cdot \frac{1}{\left|1 + \sum_{j=1}^{p} \alpha_{jp} e^{ijx}\right|^2} = \frac{1}{2\pi} \cdot \frac{1}{\left|\varphi_p^*(e^{ix})\right|^2}$$

Also, $f(x) = \frac{1}{2\pi} \cdot \frac{1}{\left|\pi(e^{ix})\right|^2}$ a.e. (theorem 3.4)

Theorem 4.1:

Under the condition (1.1), we have that

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \cdot f(x)\, dx = 0 \tag{1.2}$$

and

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \frac{|f(x) - f_p(x)|}{f_p(x)}\, dx = 0 \tag{1.3}$$

Proof:

Note first that

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| = \left| \frac{f(x) - f_p(x)}{f(x) \cdot f_p(x)} \right|$$

So (1.2) implies (1.3).

Now,

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| = 2\pi \left| |\pi(e^{ix})|^2 - |\varphi_p^*(e^{ix})|^2 \right|$$

$$\leq 2\pi \left(|\pi(e^{ix})| + |\varphi_p^*(e^{ix})| \right) \cdot \left| \pi(e^{ix}) - \varphi_p^*(e^{ix}) \right| .$$

Then

$$\int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \cdot f(x)\, dx \leq 2\pi \int_{-\pi}^{\pi} \left(|\pi(e^{ix})| + |\varphi_p^*(e^{ix})| \right) \cdot |\pi(e^{ix}) - \varphi_p^*(e^{ix})| \cdot f(x)\, dx$$

Using Schwarz' inequality, we obtain

$$\int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \cdot f(x)\, dx \leq 2\pi \left(\|\pi(e^{ix})\|_F + \|\varphi_p^*(e^{ix})\|_F \right) \cdot \|\pi(e^{ix}) - \varphi_p^*(e^{ix})\|_F$$

Recall that $\|\mu(\cdot)\|_F = \left(\int_{-\pi}^{\pi} |\mu(e^{ix})|^2 \cdot f(x)\, dx \right)^{\frac{1}{2}}$.

We have shown previously that

$$\|\varphi_p^*(e^{ix})\|_F = 1 \qquad (3.2.1.16)$$

and by Theorem 3.4

$$\|\pi(e^{ix})\|_F = \sqrt{\sum_{j=0}^{\infty} |\varphi_j(0)|^2} \Big/ \sqrt{K(0,0)} = 1 \ ,$$

$$\|\pi(e^{ix}) - \varphi_p^*(e^{ix})\|_F \xrightarrow[p \to \infty]{} 0 \ ,$$

which completes the proof.

Note that (1.2) and all other expressions involving $\frac{1}{f(\cdot)}$ can be rewritten using the special notation introduced in subsection 3.1.3, e.g. for (1.2)

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \frac{|fi(x) - fi_p(x)|}{fi(x)}\, dx = 0 \quad .$$

Thus the statements about $f(\cdot)$ are not exactly of the same form as the statements about $fi(\cdot)$.

For our second step we add some conditions to insure that the orthogonal polynomials $\{\varphi_n(\cdot)\}_{n=0}^{\infty}$ are uniformly bounded and so also the $\{\varphi_n^*(\cdot)\}_{n=0}^{\infty}$.

Recall that $f(\cdot) \in \text{Lip}(\alpha,2)$ if

$$w_2(\delta;f) = \sup_{|h|\le\delta} \left[\int_{-\pi}^{\pi} |f(x+h) - f(x)|^2\, dx\right]^{\frac{1}{2}} = O(\delta^{\alpha}) \quad .$$

(1.4)
$$\begin{cases} \text{condition (1.1)} \\ 0 < m \le f(x) \le M < \infty \quad , \quad \text{a.e. in } [-\pi,\pi] \\ f(\cdot) \in \text{Lip}(1/2,2) \end{cases}$$

Theorem 4.2:

Under condition (1.4),

(1.5)
$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \left|\frac{1}{f(x)} - \frac{1}{f_p(x)}\right|^2 dx = 0 \quad ,$$

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \left| \frac{f(x) - f_p(x)}{f_p(x)} \right|^2 dx = 0 \tag{1.6}$$

Proof:

Note first that

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| = \frac{|f(x) - f_p(x)|}{f(x) \cdot f_p(x)} \geq \frac{1}{M} \frac{|f(x) - f_p(x)|}{f_p(x)}$$

and so (1.5) implies (1.6).

Again

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \leq 2\pi \left(|\pi(e^{ix})| + |\varphi_p^*(e^{ix})| \right) \cdot \left| \pi(e^{ix}) - \varphi_p^*(e^{ix}) \right|$$

But under condition (1.4),

$$|\pi(e^{ix})| \leq \frac{1}{\sqrt{m}} \quad , \qquad f(\cdot) \text{ being bounded below}$$

and

$$|\varphi_p^*(e^{ix})| \leq C \quad , \qquad \text{by Theorem 3.6} \quad .$$

So

$$m \cdot \int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right|^2 dx \leq \int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right|^2 \cdot f(x)\, dx$$

$$\leq C' \cdot \int_{-\pi}^{\pi} \left| \pi(e^{ix}) - \varphi_p^*(e^{ix}) \right|^2 \cdot f(x)\, dx$$

which goes to zero according to Theorem 3.4.

This completes the proof.

A slight modification of condition (1.4) will provide us with pointwise convergence.

4.1.2 Pointwise convergence

$$(2.1)\qquad \begin{cases} \text{condition (1.1)} \\ 0 < m \le f(x) \le M < \infty \quad , \quad \text{a.e. in } [-\pi,\pi] \\ f(\cdot) = g(\cdot) \quad , \quad \text{a.e. in } [-\pi,\pi] \\ w_2(\delta;g) = O(\delta^{\alpha}) \quad , \quad \frac{1}{2} < \alpha \le 1 \end{cases}$$

Theorem 4.3:

Under condition (2.1),

$$(2.2)\qquad \lim_{p\to\infty} \int_{-\pi}^{\pi} |f(x) - f_p(x)|^2 \, dx = 0$$

and

$$\lim_{p\to\infty} f_p(x) = \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2} \quad , \quad \text{uniformly}$$

$$\left(f(x) = \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2} \quad , \quad \text{a.e.}\right)$$

Proof:

Under condition (2.1), we certainly have (1.4); but by Lemma 5 of Ibragimov (1964), $\{|\varphi_n(\cdot)|\}_{n=0}^{\infty}$ is uniformly bounded from above and

below, i.e. $0 < b \le |\varphi_n(e^{ix})| \le B < \infty$. Thus

$$0 < a \le f_p(x) \le A < \infty \quad , \quad \text{for all } x \text{ and all } p$$

and (2.2) follows from (1.6).

To prove (2.3), we note that Theorem 3.7 can be applied so that

$$\lim_{p\to\infty} \varphi_p^*(e^{ix}) = \pi(e^{ix}) \quad , \quad \text{uniformly}$$

and by the previous remark ($\{|\varphi_p^*(\cdot)|\}_{p=0}^{\infty}$ is uniformly bounded from above and below)

$$\frac{1}{2\pi} \lim_{p\to\infty} \frac{1}{|\varphi_p^*(e^{ix})|^2} = \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2} \quad , \quad \text{uniformly} .$$

This completes the proof.

At what rate does the bias decrease to zero?

(2.4)
$$\begin{cases} \text{condition (1.1)} & \\ 0 < m \le f(\ ) \le M < \infty \quad , & \text{a.e. in } [-\pi,\pi] \\ f(\cdot) = g(\cdot) \quad , & \text{a.e. in } [-\pi,\pi] \\ g(\cdot) \text{ has } r \text{ derivatives} & \\ g^{(r)}(\cdot) \in \text{Lip}(\alpha,2) \quad , & 0 < \alpha \le 1 \end{cases}$$

Theorem 4.4:

Under condition (2.1),

$$|b_p(x)| = O(p^{-\beta}) \quad , \qquad \beta < \alpha - 1/2 \quad .$$

Under condition (2.4),

$$|b_p(x)| = O(p^{-\beta}) \quad , \qquad \beta < r + \alpha - 1/2 \quad .$$

Proof:

This is essentially Theorem 3.12 in Kromer (1969).

4.1.3 Properties based on the partial correlation coefficients

As in Section 3.2.2, Part C, we can use the partial correlation coefficients to describe the properties of the bias function. We obtain results that are similar to those of the two previous sections.

Theorem 4.5:

$$\text{If} \quad \sum_{n=1}^{\infty} |\alpha_{nn}|^2 < \infty \quad ,$$

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \cdot f(x)\, dx = 0$$

and

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} \left| \frac{f(x) - f_p(x)}{f_p(x)} \right| dx = 0$$

Proof: By Theorem 3.9, $\sum_{n=1}^{\infty} |\alpha_{nn}|^2 < \infty$ implies that $\log f(\cdot)$ is Lebesgue-integrable. So we can apply Theorem 4.1.

Theorem 4.6:

If $|\alpha_{nn}| < \frac{3}{n \log n}$, for n large enough,

$$\lim_{p\to\infty} f_p(x) = f(x) \quad , \quad \text{at all } x \text{ such that } 0 < f(x) .$$

Proof: By Theorem 3.10, we have that

$$\lim_{p\to\infty} \varphi_p^*(e^{ix}) = \pi(e^{ix}) \quad , \text{ at all } x \text{ where } f(x) > 0$$

and so

$$\lim_{p\to\infty} \frac{1}{|\varphi_p^*(e^{ix})|^2} = \frac{1}{|\pi(e^{ix})|^2} .$$

We can get now different estimates of the bias.

Theorem 4.7:

If $\sum_{n=1}^{\infty} |\alpha_{nn}| < \infty$, then, $F(\cdot)$ is absolutely continuous

$$f(x) \geq m > 0 \quad , \qquad x \in [-\pi,\pi]$$

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \leq C \cdot \sum_{k=p}^{\infty} |\alpha_{kk}|$$

Proof:

The first two assertions follow from Theorem 3.12. Then, as in Theorem 4.2,

$$\left| \frac{1}{f(x)} - \frac{1}{f_p(x)} \right| \leq 2\pi \cdot \left(\frac{1}{\sqrt{m}} + C' \right) \cdot |\pi(e^{ix}) - \varphi_p^*(e^{ix})|$$

$$\leq C \cdot \sum_{k=p}^{\infty} |\alpha_{kk}| \quad , \quad \text{by Theorem 3.12.}$$

If we add to the hypothesis of Theorem 4.7 that $f(\cdot)$ is bounded from above, then we can prove that $|f(x) - f_p(x)|$ goes to zero uniformly at the same rate as $\sum_{k=p}^{\infty} |\alpha_{kk}|$.

4.2 Fourier analysis

The Fourier series representation of $f(\cdot)$ is

$$f(x) \sim \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R(v)\, e^{-ivx}$$

and the Fourier series representation of $f_p(\cdot)$ is

$$f_p(x) \sim \frac{1}{2\pi} \sum_{v=-\infty}^{\infty} R_p(v)\, e^{-ivx} \quad ,$$

where

$$R_p(v) = \begin{cases} R(v) \quad , & |v| \le p \\ -\sum_{j=1}^{p} \alpha_{jp} R_p(v+j) \quad , & v < -p \\ R_p(-v) \quad , & v > p \end{cases}$$

Now it is always true that the Fourier series representation of $f_p(\cdot)$ converges pointwise to $f_p(\cdot)$ for almost all x in $[-\pi,\pi]$. This follows from the fact that $|R_p(v)|$ decreases exponentially (see Appendix 4.A.1).

If $\sum_{v=-\infty}^{\infty} |R(v)|^2 < \infty$, then $f(\cdot)$ is square-integrable and

$$\lim_{p\to\infty} \int_{-\pi}^{\pi} |f(x) - f_p(x)|^2 \, dx = 0 \quad .$$

If $\sum_{v=-\infty}^{\infty} |R(v)| < \infty$, the Fourier series of $f(\cdot)$ converges pointwise to $f(\cdot)$ and we can write the bias function exactly as

$$\left|b_p(x)\right| = \frac{1}{2\pi} \cdot \left| \sum_{|v|>p} \left(R(v) - R_p(v) \right) e^{-ivx} \right| \quad \text{a.e.}$$

We obtain the following bound

$$\left|b_p(x)\right| \leq \frac{1}{2\pi} \left(\sum_{|v|>p} |R(v)| + \sum_{|v|>p} |R_p(v)| \right) \quad .$$

And so for almost every x in $[-\pi,\pi]$, the rate of fall-off of $b_p(x)$ to zero is the slowest of the rates of convergence of $\Sigma|R(v)|$ and $\Sigma|R_p(v)|$.

Appendix

4.A.1 Rate of fall-off of an autoregressive covariance function.

Suppose $\{X(t), t \in \mathbb{Z}\}$ is a stationary autoregressive process of order p, i.e., there exists an orthogonal process $\{\epsilon(t), t \in \mathbb{Z}\}$ such that

$$X(t) + \sum_{j=1}^{p} \alpha_{jp} X(t-j) = \epsilon(t) \quad ,$$

then

$$Y(t) = \Big(X(t), X(t-1), \ldots, X(t-p+1)\Big)^T$$

is a Markov process and as such its covariance function is of the form

$$R_Y(v) = R_Y(0) \cdot [\rho_Y(1)]^v \quad .$$

Now

$$R_Y(v) = \begin{bmatrix} R(v) & \cdots & R(v-p+1) \\ \vdots & & \vdots \\ R(v+p-1) & \cdots & R(v) \end{bmatrix} \quad , \text{ where } \quad R(v) = E[X(t)\,\overline{X(t+v)}]$$

and

$$\rho_Y(1) = \begin{bmatrix} -\bar{\alpha}_{1p} & 1\ 0 & \cdots & 0 \\ -\bar{\alpha}_{2p} & 0\ 1\ 0 & \cdots & 0 \\ \vdots & & & \\ -\bar{\alpha}_{p-1,p} & 0\ 0 & \cdots & 1 \\ -\bar{\alpha}_{pp} & 0\ 0 & \cdots & 0 \end{bmatrix}$$

The characteristic polynomial of $\rho_Y(1)$ is simply

$$\psi(\lambda) = \lambda^p + \sum_{j=1}^{p} \bar{\alpha}_{jp} \lambda^{p-j}$$

which has also been referred to as the indicial polynomial of the constants $\{1,\bar{\alpha}_{1p},\ldots,\bar{\alpha}_{pp}\}$. Pagano (1972) has shown that the stationarity of $X(\cdot)$ implies that all the zeros of $\psi(\cdot)$ are strictly within the unit circle.

Thus all the eigenvalues of $\rho_Y(1)$ have modulus less than 1.

By using the Jordan canonical decomposition of $\rho_Y(1)$, it is seen that

$$\lim_{v\to\infty} [\rho_Y(1)]^v = 0_{pxp}$$

and by the same token that $R(v)$ goes to zero exponentially.

CHAPTER 5

CONSISTENCY STUDY

The Autoregressive Method as an Estimation Method

In this chapter, we want to establish the consistency of the autoregressive method and find the rate of consistency in terms of a relation between the order of the estimator and the sample size.

5.1 Convergence of $\hat{R}(\cdot)$ to $R(\cdot)$

The autoregressive estimator of order p depends on the data through the $\hat{R}(\cdot)$ subsequence and through the extreme values as the data is rescaled to $[-\pi,\pi]$. Typically

$$\hat{R}(v) = \int_{-\pi}^{\pi} e^{ivx}\, dF_n(x)$$

where $F_n(x)$ is usually a step-function and, in many statistical applications, $F_n(x)$ is a function of the empirical distribution function.

$\{F_n(x),\ x \in [-\pi,\pi]\}$ is a stochastic process indexed by x. We assume throughout this chapter that this process has the following properties:

(1.1)
$\{\sqrt{n}(F_n(x) - F(x)),\ x\in[-\pi,\pi]\}$ converges to a Gaussian process with mean 0 and covariance function $\sigma(x,y)$, $x,y \in [-\pi,\pi]$, where

- $F(\cdot)$ has a derivative $f(\cdot) \geq 0$, integrable
- $-\infty < \int_{-\pi}^{\pi} \log f(x)\, dx < \infty$
- $\int_{-\pi}^{\pi} \frac{1}{f(x)}\, dx < \infty$
- $-\infty < \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sigma(x,y)\, dx\, dy < \infty$
- $\int_{-\pi}^{\pi} E[F_n(x)]\, dx \to \int_{-\pi}^{\pi} F(x)\, dx$
- $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} n\, \mathrm{Cov}\left(F_n(x),F_n(y)\right) dx\, dy \to \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sigma(x,y)\, dx\, dy$

Theorem 5.1:

Under condition (1.1), for any $p \geq 1$,

$$(1.2)\quad \sqrt{n}\begin{pmatrix} \mathrm{Re}(\hat{R}(v) - R(v)) \\ \mathrm{Im}(\hat{R}(v) - R(v)) \end{pmatrix} \xrightarrow[n\to\infty]{D} \mathrm{MVN}\big(Q, S(v)\big), \qquad |v| \leq p$$

where

$$S(v) = \begin{bmatrix} A(v) & B(v) \\ B(v) & C(v) \end{bmatrix}$$

$$A(v) = A(0) + 2 \cdot (-1)^v \int_{-\pi}^{\pi} v \sin vx \big(\sigma(\pi,x) - \sigma(-\pi,x)\big)\, dx + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} v^2 \sin vx \sin vy\ \sigma(x,y)\, dxdy$$

$$A(0) = \sigma(\pi,\pi) + \sigma(-\pi,-\pi) - 2\sigma(\pi,-\pi)$$

$$B(v) = (-1)^v \int_{-\pi}^{\pi} v \cos vy \big(\sigma(-\pi,y) - \sigma(\pi,y)\big)\, dy - \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} v^2 \sin vx \cos vy\ \sigma(x,y)\, dxdy$$

$$B(0) = 0$$

$$C(v) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} v^2 \cos vx \cos vy\ \sigma(x,y)\, dxdy$$

$$C(0) = 0 \quad .$$

(The mode of convergence is in distribution.)

Proof:

$$\hat{R}(v) = \int_{-\pi}^{\pi} e^{ivx}\, d F_n(x) \quad , \qquad |v| \le p \ .$$

Using integration by parts, we obtain

$$\hat{R}(v) = (-1)^v [F_n(\pi) - F_n(-\pi) \quad \int_{-\pi}^{\pi} ive^{ivx} F_n(x)\, dx \quad . \tag{1.3}$$

Taking expectation of both sides,

$$E[\hat{R}(v)] \underset{n\to\infty}{\longrightarrow} (-1)^v [F(\pi) - F(-\pi)] - \int_{-\pi}^{\pi} ive^{ivx} F(x)\, dx$$

which is equal to $R(v)$ as can be seen by integrating by parts again.

For the covariance, we verify the exactitude of $B(v)$ only; the other terms can be found in the same way.

$$n\, \mathrm{Cov}\left(\mathrm{Re}\left(\hat{R}(v) - R(v)\right) , \ \mathrm{Im}\left(\hat{R}(v) - R(v)\right)\right)$$
$$= n\left[E \int_{-\pi}^{\pi} \cos vx\, d\left(F_n(x) - F(x)\right) \cdot \int_{-\pi}^{\pi} \sin vy\, d\left(F_n(y) - F(y)\right)\right] \quad .$$

Integrating by parts, we obtain

$$n\, E\left[\left\{(-1)^v\left(F_n(\pi) - F_n(\pi)\right| - (-1)^v\left(F_n(-\pi) - F(-\pi)\right) + \int_{-\pi}^{\pi} v \sin vx \left(F_n(x) - F(x)\right) dx\right\}\right.$$
$$\left. \cdot \left\{-\int_{-\pi}^{\pi} v \cos vy \left|F_n(y) - F(y)\right|\, dy\right\}\right] \quad .$$

It is easy to obtain $B(v)$ from that point.

Finally, as the integral of a Gaussian process is Gaussian, we obtain the third element of our theorem, i.e., asymptotic normality. This completes the proof.

It is also clear that $\hat{R}(v)$ is a consistent estimate of $R(v)$, that is

$$\hat{R}(v) \xrightarrow[n\to\infty]{P} R(v) \quad , \quad |v| \le p \quad , \quad p \ge 1 \quad . \tag{1.4}$$

This follows from Theorem 5.1 and Chebyshev's inequality.

We can then get a multivariate analogue of Theorem 5.1.

Theorem 5.2:

Under condition (1.1),

$$\sqrt{n}\begin{pmatrix} \begin{pmatrix} \mathrm{Re}\left(\hat{R}(0) - R(0)\right) \\ \mathrm{Im}\left(\hat{R}(0) - R(0)\right) \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \mathrm{Re}\left(\hat{R}(k) - R(k)\right) \\ \mathrm{Im}\left(\hat{R}(k) - R(k)\right) \end{pmatrix} \end{pmatrix} \xrightarrow[n\to\infty]{D} \mathrm{MVN}\left(Q, \Sigma(k)\right) \quad , \quad k \le p$$

where

$$\Sigma(k) = \left[\Sigma_{j\ell} \right], \qquad j,\ell = 0,1,\ldots,k$$

$$\Sigma_{jj} = S(j) \qquad \text{defined in Theorem 5.1}$$

$$\Sigma_{j\ell} = \begin{bmatrix} A_{j\ell} & B_{j\ell} \\ D_{j\ell} & C_{j\ell} \end{bmatrix} \qquad j < \ell$$

$$\Sigma_{\ell j} = \Sigma_{j\ell}$$

$$A_{j\ell} = (-1)^{j+\ell} A(0) + (-1)^{j} \int_{-\pi}^{\pi} \ell \sin \ell y (\sigma(\pi,y) - \sigma(-\pi,y)) \, dy$$

$$+ (-1)^{\ell} \int_{-\pi}^{\pi} j \sin jy (\sigma(\pi,y) - \sigma(-\pi,y)) \, dy$$

$$+ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} j\ell \sin jx \sin \ell y \, \sigma(x,y) \, dx \, dy$$

$$A_{0\ell} = A(0) \quad , \qquad \text{as in Theorem 5.1}$$

$$B_{j\ell} = (-1)^{j} \int_{-\pi}^{\pi} \ell \cos \ell y (\sigma(-\pi,y) - \sigma(\pi,y)) \, dy$$

$$- \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} j\ell \sin jx \cos \ell y \, \sigma(x,y) \, dx \, dy$$

$$D_{j\ell} = (-1)^{\ell} \int_{-\pi}^{\pi} j \cos jy \, (\sigma(-\pi,y) - \sigma(\pi,y)) \, dy$$

$$- \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ell j \sin \ell x \cos jy \, \sigma(x,y) \, dx \, dy$$

$$C_{j\ell} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} j\ell \cos jx \cos \ell y \, \sigma(x,y) \, dx \, dy$$

$$C_{0\ell} = B_{0\ell} = 0 \quad .$$

Proof:

The proof is really equivalent to that of Theorem 5.1 and so it is omitted.

We will come back to these formulas in the next chapter when we examine three different routes to density estimation.

5.2 Consistency of $\hat{\underset{\sim}{\alpha}}_p$ and $\hat{K}_p$

It would be possible without a doubt to carry out the same program as Kromer (1969). But our main interest lies in relating the order of the autoregressive estimator to the sample size.

Lemma 5.3:

For any p-dimensional vector $\underset{\sim}{x}_p$ and matrix X_p (random or not), define

$$\|\underset{\sim}{x}_p\|^2 = \sum_{i=1}^{p} |x_i|^2$$

$$\|X_p\|^2 = \sum_{i,j=1}^{p} |X_{ij}|^2$$

$$\|X_p\|_H = \sup_{\|\underset{\sim}{x}_p\|=1} \{\|X_p \cdot \underset{\sim}{x}_p\|\} \quad .$$

Then,

$$\|X_p \cdot \underset{\sim}{x}_p\| \leq \|X_p\|_H \cdot \|\underset{\sim}{x}_p\| \tag{2.1}$$

$$\begin{cases} \|X_p\|_H \leq \|X_p\| \leq p \cdot \max\{|X_{ij}|, i,j = 1,\ldots,p\} \\ \|\underset{\sim}{x}_p\| \leq \sqrt{p} \cdot \max\{|x_i|, i = 1,\ldots,p\} \end{cases} \tag{2.2}$$

(2.3) If X_p is Hermitian and positive definite, then

$$\|X_p\|_H = \lambda_{max}(X_p)$$

$$\|X_p^{-1}\|_H = \frac{1}{\lambda_{min}(X_p)}$$

where $\lambda_{max}(X_p)$ ($\lambda_{min}(X_p)$) is the maximum (minimum) eigenvalue of X_p.

(2.4) If Y_p is non-singular and

$$\|X_p - Y_p\|_H \leq \frac{1-\epsilon}{\|Y_p^{-1}\|_H} , \qquad \epsilon > 0$$

then, $$\|X_p^{-1}\|_H \leq \frac{\|Y_p^{-1}\|_H}{\epsilon} .$$

Proof:

Most of these assertions are well-known. For (2.4), see Davies (1973).

Theorem 5.4:

Let $f(\cdot)$ be integrable and

$$m \leq f(x) \leq M , \qquad \text{a.e. in } [-\pi,\pi] .$$

Let $R(v) = \int_{-\pi}^{\pi} e^{ivx} f(x)\, dx$,

Let

$$R_p = \begin{bmatrix} R(0) & \cdots & R(p-1) \\ \vdots & & \vdots \\ R(1-p) & \cdots & R(0) \end{bmatrix}$$

Then

$$(2.5) \qquad 2\pi m \leq \lambda_{min}(R_p) \leq \lambda_{max}(R_p) \leq M \cdot 2\pi$$

and

$$(2.6) \qquad \begin{cases} \lim_{p\to\infty} \lambda_{min}(R_p) = m \cdot 2\pi \\ \lim_{p\to\infty} \lambda_{max}(R_p) = M \cdot 2\pi \end{cases}$$

Proof:

See Grenander and Szegö (1958), Chapter 5.

Theorem 5.5:

If $0 < m \leq f(x) \leq M < \infty$, a.e. in $[-\pi,\pi]$ and

$$\lim_{n\to\infty} \frac{p(n)}{\sqrt{n}} = 0 \quad ,$$

then

$$(2.7)\qquad \begin{cases} \| \hat{\underset{\sim}{\alpha}}_{p(n)} - \underset{\sim}{\alpha}_{p(n)} \| \xrightarrow[n\to\infty]{P} 0 \\ \| \hat{K}_{p(n)} - K_{p(n)} \| \xrightarrow[n\to\infty]{P} 0 \end{cases}$$

Note: From now on, we will use p instead of $p(n)$ to simplify the notation, but with the understanding that p is allowed to increase with n at the rate specified.

Proof:

Let

$$\underset{\sim}{\alpha}_p = (\alpha_{1p},\ldots,\alpha_{pp})'$$

$$\hat{\underset{\sim}{\alpha}}_p = (\hat{\alpha}_{1p},\ldots,\hat{\alpha}_{pp})'$$

$$\underset{\sim}{r}_p = (R(-1),\ldots,R(-p))'$$

$$\hat{\underset{\sim}{r}}_p = (\hat{R}(-1),\ldots,\hat{R}(-p))'$$

$$R_p = \begin{bmatrix} R(0) & \cdots & R(p-1) \\ \vdots & & \vdots \\ R(1-p) & \cdots & R(0) \end{bmatrix}$$

$$\hat{R}_p = \begin{bmatrix} \hat{R}(0) & \cdots & \hat{R}(p-1) \\ \vdots & & \vdots \\ \hat{R}(1-p) & \cdots & \hat{R}(0) \end{bmatrix}$$

We then have

$$(2.8) \qquad \underset{\sim}{\alpha}_p - \hat{\underset{\sim}{\alpha}}_p = \hat{R}_p^{-1}\left[(\hat{\underset{\sim}{r}}_p - \underset{\sim}{r}_p) + (\hat{R}_p - R_p)\,\underset{\sim}{\alpha}_p\right] \quad .$$

This is true because the Yule-Walker equations can be written

$$R_p \underset{\sim}{\alpha}_p = -\underset{\sim}{r}_p$$

$$\hat{R}_p \hat{\underset{\sim}{\alpha}}_p = -\hat{\underset{\sim}{r}}_p \quad .$$

$$\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| \le \|\hat{R}_p^{-1}\|_H \cdot \left[\|\hat{\underset{\sim}{r}}_p - \underset{\sim}{r}_p\| + \|\hat{R}_p - R_p\| \cdot \|\underset{\sim}{\alpha}_p\|\right] \text{, by (2.1)} .$$

Now, $\|\hat{R}_p - R_p\|_H \le p \cdot \max\{|\hat{R}(v) - R(v)| \;,\; |v| \le p\}$.

By Theorem 5.2,

$$(2.9) \qquad |\hat{R}(v) - R(v)| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

meaning that $\hat{R}(v)$ converges in probability to $R(v)$ independently of v at the rate of $1/\sqrt{n}$ as $n \to \infty$.

Thus

$$\|\hat{R}_p - R_p\| = O_p\left(\frac{p}{\sqrt{n}}\right) .$$

Now as $\lim_{n\to\infty} \frac{p}{\sqrt{n}} = 0$, there exists $\epsilon > 0$ such that

$$\frac{p}{\sqrt{n}} < \frac{1-\epsilon}{\|R_p^{-1}\|_H} = (1-\epsilon) \cdot \lambda_{min}(R_p) \text{ , by (2.3)}$$

$$\leq (1-\epsilon) \cdot M \cdot 2\pi \text{ by Theorem 5.4 .}$$

$$\text{So, } \|\hat{R}_p^{-1}\|_H \leq \frac{\|R_p^{-1}\|_H}{\epsilon} \text{ , by (2.4)}$$

$$\leq \frac{1}{\epsilon \cdot \lambda_{min}(R_p)} \text{ , by (2.3)}$$

$$\leq \frac{1}{\epsilon \cdot m \cdot 2\pi} \text{ , by Theorem 5.4 .}$$

Thus,

$$\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| \leq \frac{1}{\epsilon \cdot m \cdot 2\pi}\left[\|\hat{\underset{\sim}{r}}_p - \underset{\sim}{r}_p\| + \|\hat{R}_p - R_p\| \cdot \|\underset{\sim}{\alpha}_p\|\right] ;$$

$$\|\underset{\sim}{\alpha}_p\| = \sqrt{\sum_{j=1}^{p} |\alpha_{jp}|^2} \xrightarrow[p \to \infty]{} \sqrt{\sum_{j=1}^{\infty} |\alpha_j|^2} < \infty \text{ , by Theorem 3.3 ;}$$

$$\|\hat{\underset{\sim}{r}}_p - \underset{\sim}{r}_p\| \leq \sqrt{p} \cdot \max\{|\hat{R}(v) - R(v)| , |v| \leq p\} \text{ , by (2.2)}$$

$$= o_p\left(\frac{\sqrt{p}}{\sqrt{n}}\right) \text{ , by (2.9) ;}$$

$$\|\hat{R}_p - R_p\| = o_p\left(\frac{p}{\sqrt{n}}\right) \text{ , by (2.2) and (2.9) .}$$

So

$$\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| = o_p\left(\frac{p}{\sqrt{n}}\right)$$

$$\text{i.e.,} \quad \|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| \xrightarrow[n \to \infty]{P} 0 \quad , \quad \text{if} \quad \frac{p}{\sqrt{n}} \xrightarrow[n \to \infty]{} 0 \quad .$$

For the second part of the theorem,

$$\hat{K}_p - K_p = \sum_{j=0}^{p} \hat{\alpha}_{jp} \hat{R}(j) - \sum_{j=0}^{p} \alpha_{jp} R(j) \quad , \quad \text{by}$$

$$= \sum_{j=0}^{p} \{\hat{\alpha}_{jp}(\hat{R}(j) - R(j)) + R(j)\,(\hat{\alpha}_{jp} - \alpha_{jp})\} .$$

$$|\hat{K}_p - K_p| \le \sum_{j=0}^{p} \{|\hat{\alpha}_{jp}| \cdot |\hat{R}(j) - R(j)| + |R(j) \cdot |\hat{\alpha}_{jp} - \alpha_{jp}|\}$$

$$\le \left(\sum_{j=0}^{p} |\hat{\alpha}_{jp}|^2\right)^{\frac{1}{2}} \cdot \sqrt{p} \cdot \max\{|\hat{R}(j) - R(j)| \, , \, |j| \le p\}$$

$$+ \left(\sum_{j=0}^{p} |R(j)|^2\right)^{\frac{1}{2}} \cdot \|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\|$$

$$\le \left(1 + \|\underset{\sim}{\alpha}_p\|\right) \cdot o_p\left(\frac{\sqrt{p}}{\sqrt{n}}\right) + \left(\sum_{j=0}^{p} |R(j)|^2\right)^{\frac{1}{2}} \cdot o_p\left(\frac{p}{\sqrt{n}}\right) .$$

Now, we certainly have

$$\int_{-\pi}^{\pi} |f(x)|^2 \, dx \le 2\pi \cdot M^2 < \infty$$

and so

$$\left(\sum_{j=0}^{p} |R(j)|^2 \right)^{\frac{1}{2}} \xrightarrow[p \to \infty]{} \left(\sum_{j=0}^{\infty} |R(j)|^2 \right)^{\frac{1}{2}} < \infty$$

($\{R(j)\}_{j=0}^{\infty}$ is the set of Fourier coefficients of $f(\cdot)$). Finally, as

$$\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| \xrightarrow{P} 0$$

$$\|\hat{\underset{\sim}{\alpha}}_p\| \xrightarrow{P} \left(\sum_{j=1}^{\infty} |\alpha_j|^2 \right)^{\frac{1}{2}} < \infty$$

so, $|\hat{K}_p - K_p| = O_p\left(\frac{p}{\sqrt{n}}\right)$.

This completes the proof.

The $O_p(\cdot)$ notation was introduced by Mann and Wald (1943). Redefine $\hat{\underset{\sim}{\alpha}}_p = (\hat{\alpha}_{op},\ldots,\hat{\alpha}_{pp})'$ and $\underset{\sim}{\alpha}_p = (\alpha_{op},\ldots,\alpha_{pp})'$. Theorem 5.5 still holds because $\hat{\alpha}_{op} = \alpha_{op} = 1$, and it follows from it that $\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}\| \xrightarrow{P} 0$, where $\alpha = (\alpha_0,\alpha_1,\alpha_2,\ldots)'$, because

$$\|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}\| \leq \|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\| + \|\underset{\sim}{\alpha}_p - \underset{\sim}{\alpha}\|$$

and

$$\|\underset{\sim}{\alpha}_p - \underset{\sim}{\alpha}\|^2 = \int_{-\pi}^{\pi} \left| \sum_{j=0}^{\infty} (\alpha_{jp} - \alpha_j)\, e^{ijx} \right|^2 \cdot dx \ , \quad (\alpha_{jp} = 0 \ , \ j > p)$$

$$\leq \frac{1}{m} \cdot \int_{-\pi}^{\pi} \left| \sum (\alpha_{jp} - \alpha_j)\, e^{ijx} \right|^2 \cdot f(x)\, dx$$

$$\leq \frac{K}{m} \cdot \left[\|\pi(e^{ix}) - \varphi_p^*(e^{ix})\|_F + \left| 1 - \frac{\sqrt{K_p}}{\sqrt{K}} \right| \cdot \|\varphi_p^*(e^{ix})\|_F \right]^2 .$$

$\xrightarrow[p \to \infty]{} 0$, by Theorem 3.4, (3.3.0.12) and (3.2.1.16).

Also, $\|\hat{K}_p - K\| \xrightarrow{P} 0$, as $K_p \to K$ by (3.3.0.12).

Berk (1974) also proves the consistency of the autoregressive coefficients, but his rate of consistency is p^3/n and he has to make the further assumption that

$$p \cdot \sum_{j=p+1}^{\infty} |a_j|^2 \xrightarrow[p\to\infty]{} 0$$

(see his Lemma 3 and equation 2.17).
We only require $p^2/n \to 0$. Even though our contexts are different, our proof can be adapted to Berk's context.

5.3 Consistency of $\hat{f}_p(\cdot)$

From Theorem 5.5, we will now obtain results that parallel those of Chapter 4.

Lemma 5.6:

If $0 < m \le f(x) \le \infty$, a.e., in $[-\pi,\pi]$ and $\lim_{n\to\infty} \frac{p^{3/2}}{\sqrt{n}} = 0$, then

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| \xrightarrow[n\to\infty]{P} 0 \quad , \text{ uniformly in } x .$$

Proof:

$$\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix}) = \frac{1}{\sqrt{\hat{K}_p}} \sum_{j=0}^{p} \hat{\alpha}_{jp} e^{ijx} - \frac{1}{\sqrt{K_p}} \sum_{j=0}^{p} \alpha_{jp} e^{ijx}$$

$$= \frac{1}{\sqrt{\hat{K}_p}} \sum_{j=0}^{p} (\hat{\alpha}_{jp} - \alpha_{jp}) e^{ijx} + \left(\frac{1}{\sqrt{\hat{K}_p}} - \frac{1}{\sqrt{K_p}} \right) \sum_{j=0}^{p} \alpha_{jp} e^{ijx} .$$

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| \le \frac{\sum_{j=0}^{p} |\hat{\alpha}_{jp} - \alpha_{jp}|}{\sqrt{\hat{K}_p}} + \left| \frac{1}{\sqrt{\hat{K}_p}} - \frac{1}{\sqrt{K_p}} \right| \sum_{j=0}^{p} |\alpha_{jp}|$$

$$\le \frac{\sqrt{p} \cdot \|\hat{\underset{\sim}{\alpha}}_p - \underset{\sim}{\alpha}_p\|}{\sqrt{\hat{K}_p}} + \left| \frac{1}{\sqrt{\hat{K}_p}} - \frac{1}{\sqrt{K_p}} \right| \cdot \sqrt{p}\, \|\underset{\sim}{\alpha}_p\| .$$

$$\text{As} \quad |\hat{K}_p - K_p| \xrightarrow{P} 0 \quad , \qquad |\sqrt{\hat{K}_p} - \sqrt{K_p}| \xrightarrow{P} 0$$

because $\sqrt{\cdot}$ is a continuous function and $\left|\frac{1}{\sqrt{\hat{K}_p}} - \frac{1}{\sqrt{K_p}}\right| \xrightarrow{P} 0$

because the reciprocal is a continuous function and

$0 < \sqrt{K} < \sqrt{K_p} < \sqrt{K_1} < \infty$, by (3.3.0.12) . Thus

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| = O_p\left(\frac{p^{3/2}}{\sqrt{n}}\right) \quad .$$

This completes the proof.

Theorem 5.7:

If $0 < m \le f(x) \le M < \infty$, a.e., in $[-\pi,\pi]$ and $\lim_{n\to\infty} \frac{p^{3/2}}{\sqrt{n}} = 0$, then

$$\int_{-\pi}^{\pi} \left|\frac{1}{\hat{f}_p(x)} - \frac{1}{f(x)}\right| dx \xrightarrow[n\to\infty]{P} 0$$

Proof:

$$\int_{-\pi}^{\pi} \left|\frac{1}{\hat{f}_p(x)} - \frac{1}{f(x)}\right| dx \le \int_{-\pi}^{\pi} \left|\frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)}\right| dx + \int_{-\pi}^{\pi} \left|\frac{1}{f_p(x)} - \frac{1}{f(x)}\right| dx$$

By Theorem 4.1, with the addition that $f(\cdot)$ is essentially bounded, we have that

$$\int_{-\pi}^{\pi} \left|\frac{1}{f_p(x)} - \frac{1}{f(x)}\right| dx \xrightarrow[p\to\infty]{} 0 .$$

On the other hand,

$$m \cdot \int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right| dx$$

$$\leq 2\pi \int_{-\pi}^{\pi} (|\hat{\varphi}_p^*(e^{ix})| + |\varphi_p^*(e^{ix})|) \cdot |\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| \cdot f(x)\, dx$$

$$\leq 2\pi \left[\left(\int_{-\pi}^{\pi} |\hat{\varphi}_p^*(e^{ix})|^2 \cdot f(x)\, dx \right)^{\frac{1}{2}} + \left(\int_{-\pi}^{\pi} |\varphi_p^*(e^{ix})| \cdot f(x)\, dx \right)^{\frac{1}{2}} \right]$$

$$\cdot \left[\int |\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})|^2 \cdot f(x)\, dx \right]^{\frac{1}{2}}$$

As

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| \xrightarrow{P} 0, \quad \text{uniformly in } x,$$

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})|^2 \xrightarrow{P} 0, \quad \text{uniformly in } x$$

and so

$$\int_{-\pi}^{\pi} |\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})|^2 \cdot f(x)\, dx \xrightarrow{P} 0 .$$

We also have that

$$\left| \left[\int |\hat{\varphi}_p^*(e^{ix})|^2 \cdot f(x)\, dx \right]^{\frac{1}{2}} - \left[\int_{-\pi}^{\pi} |\varphi_p^*(e^{ix})|^2 \cdot f(x)\, dx \right]^{\frac{1}{2}} \right| \xrightarrow{P} 0$$

because the square function and the integral are continuous transformations.

Finally, $\left(\int_{-\pi}^{\pi} |\varphi_p^*(e^{ix})|^2 \cdot f(x)\, dx\right)^{\frac{1}{2}} = 1$, by (3.2.1.16) .

Thus,

$$\int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right| dx = O_p\left(\frac{p^{3/2}}{\sqrt{n}}\right) .$$

This completes the proof.

Theorem 5.8:

Under condition (4.1.1.4) and if $\lim_{n\to\infty} \frac{p^{3/2}}{\sqrt{n}} = 0$

then

$$\int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f(x)} \right|^2 dx \xrightarrow{P} 0$$

Proof:

$$\int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f(x)} \right|^2 dx \le \int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right|^2 dx + \int_{-\pi}^{\pi} \left| \frac{1}{f_p(x)} - \frac{1}{f(x)} \right|^2 dx$$

By Theorem 4.2, we have that

$$\int_{-\pi}^{\pi} \left| \frac{1}{f_p(x)} - \frac{1}{f(x)} \right|^2 dx \xrightarrow[p\to\infty]{} 0 .$$

On the other hand,

$$m \cdot \int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right|^2 dx$$

$$\leq 4\pi^2 \int_{-\pi}^{\pi} \left(|\hat{\varphi}_p^*(e^{ix})| + |\varphi_p^*(e^{ix})| \right)^2 \cdot |\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})|^2 \cdot f(x)\, dx$$

By Theorem 3.6,

$$|\varphi_p^*(e^{ix})| \leq C \quad , \quad \text{independently of } x \text{ and } p \ .$$

By Lemma 5.6

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| \xrightarrow[n\to\infty]{P} 0$$

that is

$$\hat{\varphi}_p^*(e^{ix}) = \varphi_p^*(e^{ix}) + o_p(1) \leq C + o_p(1)$$

and

$$|\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})| = O_p\left(\frac{p^{3/2}}{\sqrt{n}}\right)$$

Thus,

$$\int_{-\pi}^{\pi} \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right|^2 dx \leq 4\pi^2 \int_{-\pi}^{\pi} \left(2C + c_p(1) \right)^2 \cdot O_p\left(\frac{p^3}{n}\right) \cdot f(x)\, dx$$

$$= O_p\left(\frac{p^3}{n}\right)$$

This completes the proof.

Theorem 5.9:

Under condition (4.1.2.1) and if

$$\lim_{n\to\infty} \frac{p^{3/2}}{\sqrt{n}} = 0 \quad ,$$

then

$$\left| \frac{1}{\hat{f}_p(x)} - 2\pi \cdot \left|\pi(e^{ix})\right|^2 \right| \xrightarrow{P} 0 \quad , \quad \text{uniformly}$$

and

$$\left| \hat{f}_p(x) - \frac{1}{2\pi} \cdot \frac{1}{\left|\pi(e^{ix})\right|^2} \right| \xrightarrow{P} 0 \quad , \quad \text{uniformly} \quad .$$

Proof:

$$\left| \frac{1}{\hat{f}_p(x)} - 2\pi \cdot \left|\pi(e^{ix})\right|^2 \right| \le \left| \frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)} \right| + \left| \frac{1}{f_p(x)} - 2\pi \cdot \left|\pi(e^{ix})\right|^2 \right| .$$

By Theorem 4.3,

$$\left| \frac{1}{f_p(x)} - 2\pi \cdot \left|\pi(e^{ix})\right|^2 \right| \xrightarrow[p\to\infty]{} 0 \quad , \quad \text{uniformly} \quad .$$

On the other hand,

$$\left|\frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)}\right| \leq 2\pi\left(|\hat{\varphi}_p^*(e^{ix}) + \varphi_p^*(e^{ix})|\right) \cdot |\hat{\varphi}_p^*(e^{ix}) - \varphi_p^*(e^{ix})|$$

$$\leq 2\pi\,(2C + o_p(1)) \cdot O_p\left(\frac{p^{3/2}}{\sqrt{n}}\right)$$

as in Theorem 5.8.

Thus,

$$\left|\frac{1}{\hat{f}_p(x)} - \frac{1}{f_p(x)}\right| = O_p\left(\frac{p^{3/2}}{\sqrt{n}}\right)$$

For the second part of the theorem,

$$\left|\hat{f}_p(x) - \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2}\right| \leq \left|\hat{f}_p(x) - f_p(x)\right| + \left|f_p(x) - \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2}\right|$$

By Theorem 4.3,

$$\left|f_p(x) - \frac{1}{2\pi} \cdot \frac{1}{|\pi(e^{ix})|^2}\right| \xrightarrow[p\to\infty]{} 0 \quad .$$

$$|\hat{f}_p(x) - f_p(x)| \xrightarrow[n\to\infty]{p} 0$$

because the reciprocal function is a continuous function and

$$0 < a \leq f_p(x) \leq A < \infty \quad , \quad \text{for all} \quad p \quad \text{and all} \quad x \quad .$$

This completes the proof.

The rate of consistency that is achieved is of the order of $\frac{p^3}{n}$.

CHAPTER 6

THREE WAYS TO DENSITY ESTIMATION

6.1 The Three Ways and the Basic Assumptions

We assume from the start that we want to approximate the density of a bounded (1.1) random variable whose range is taken to be $[-\pi,\pi]$ without loss of generality. Note that we can always replace the word "approximate" by "estimate." Also, we will work on the natural interval of definition of each function.

The three basic ways to approximate a density are:

— the direct approach: approximate $f(\cdot)$

— the sparsity approach: approximate $q(\cdot)$ the derivative of the quantile function and form $f(Q(t)) = \frac{1}{q(t)}$, where $Q(t) = \int_0^t q(u)\,du$

— the hazard approach: approximate $h(\cdot)$ the derivative of $(-\log(1-F(\cdot))$ and form $f(x) = h(x) \cdot \exp\left(-\int_{-\pi}^{x} h(u)\,du\right)$

The distribution function $F(\cdot)$ is always a bounded nondecreasing function and, under (1.1), the quantile function $Q(\cdot)$ is also bounded nondecreasing, but the integrated hazard $H(\cdot)$ is unbounded though nondecreasing.

Thus $H(\cdot)$ does not really fit in here even though we have obtained good empirical results in Part I of our research. It is to be noted that we never attempted to approximate $h(\cdot)$ on the whole range $[-\pi,\pi]$ but rather on $[-\pi,\pi-\epsilon]$.

The log-integrability condition can then be expressed in various ways:

$$
(1.2)\qquad \begin{cases} \displaystyle\int_{-\pi}^{\pi} \log f(x)\, dx = \int_{0}^{1} -q(t) \cdot \log q(t)\, dt \\ \displaystyle\int_{0}^{1} \log q(t)\, dt = \int_{-\pi}^{\pi} f(x) \cdot \log f(x)\, dx \\ \displaystyle\int_{-\pi}^{\pi-\epsilon} \log h(x)\, dx = \int_{-\pi}^{\pi-\epsilon} \log \frac{f(x)}{1-F(x)}\, dx \\ \displaystyle\int_{-\pi}^{\pi-\epsilon} \log h(x)\, dx = \int_{0}^{1-\delta} -q(t) \cdot \log\{(1-t) \cdot q(t)\}\, dt \end{cases}
$$

If the interval $[-\pi,\pi]$ is replaced by $(-\infty,\infty)$, then $\log q(t)$ integrable is the weakest assumption. That gives more weight to the indication that the sparsity approach might be preferred in "tough" situations as seen in Part I. Another reason to prefer the sparsity approach is that the sparsity is naturally defined on the bounded interval $[0,1]$. The only problem is then the unboundedness of the quantile function, i.e., $q(\cdot)$ will not be integrable. But research has started to extend the applicability of the autoregressive method in that direction.

6.2 The Empirical Processes

In Section 5.1 we have imposed certain conditions on the empirical process $\{F_n(x), x \in [-\pi,\pi]\}$ from which we estimate the $R(\cdot)$ sequence. We now illustrate what the general formulas look like in the case of density estimation.

6.2.1 The Direct Approach

In the direct density estimation case, $F_n(x)$ is the empirical distribution function. It is well known that $\sqrt{n}\ \{F_n(x) - F(x)\}$ converges to a Brownian Bridge process, that is to a Gaussian process with mean zero and covariance function

$$\sigma(x,y) = F(x)\big(1 - F(y)\big) , \qquad x \le y \ . \tag{1.1}$$

Using the same notation as in Section 5.1, we have for instance that

$$\begin{cases} A(v) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} v^2 \sin vx \sin vy \ \sigma(x,y)\, dx\, dy \\ B_{j\ell} = -\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} j\ell \sin jx \cos \ell y \ \sigma(x,y)\, dx\, dy \end{cases} \tag{1.2}$$

The other formulas can easily be guessed from those two.

We can simplify them even more when we carry out the integration:

$$(1.3)\qquad \begin{cases} A(v) = \mathrm{Var}(\cos vX) \\ B_{j\ell} = \mathrm{Cov}(\cos jX,\ \sin \ell X) \end{cases}$$

where X is distributed according to $F(\cdot)$. Also we find that

$$(1.4)\qquad n\ \mathrm{Cov}(\hat{R}(j),\hat{R}(\ell)) \xrightarrow[n\to\infty]{} R(j-\ell) - R(j)\,R(-\ell)$$

which can be written as

$$\int_{-\pi}^{\pi} e^{i(j-\ell)x} f(x)\,dx - \int_{-\pi}^{\pi} e^{ijx} f(x)\,dx \cdot \int_{-\pi}^{\pi} e^{-i\ell x} f(x)\,dx \quad .$$

This can be contrasted with results obtained in time series analysis as in Kromer's dissertation (1969). Let us note first that Kromer was estimating the spectral density of an observed real time series whereas we are estimating the spectral density of a complex hypothetical time series. In his Theorem 3.2, Kromer shows that

$$(1.5)\qquad n\ \mathrm{Cov}(\hat{R}(j),\hat{R}(\ell)) \longrightarrow 2\pi \cdot \int_{-\pi}^{\pi} (e^{i(j-\ell)x} + e^{i(j+\ell)x}) \cdot f^2(x)\,dx$$

where $f(\cdot)$ is the spectral density of the real time series.

We emphasize that the important difference is not between real and complex, but between observed and hypothetical. That is the difference that forces us to use a different type of estimator of the $R(\cdot)$ sequence.

So our analogy between density estimation and spectral density estimation breaks down at the point where we evaluate the variance of the estimators.

6.2.2 The Hazard Approach

In the hazard approach, the empirical process that we have used is

$$\log\left(1 - \frac{n}{n+1}\cdot F_n(x)\right) \quad , \tag{2.1}$$

where $F_n(x)$ is the empirical distribution function as in 6.2.1.

By using the delta-method, we can find its limiting distribution to be Gaussian with mean $\log(1 - F(x))$ and with covariance function

$$\sigma(x,y) = \frac{F(x)\left(1 - F(y)\right)}{\left(1 - F(x)\right)^2} \quad , \quad x \le y \quad , \tag{2.2}$$

Note that $\sigma(x,y) \xrightarrow[F(x)\to 1]{} \infty$.

The form (2.1), though satisfactory in practice, is quite unsuitable for theoretical study. We need a procedure that will never take us arbitrarily close to $F(x) = 1$.

6.2.3 The Sparsity Approach

In the sparsity approach, we use the empirical quantile process $\{Q_n(t), 0 \le t \le 1\}$. In the literature on order statistics, we find that $\sqrt{n}\,(Q_n(t) - Q(t))$ is a Gaussian process with mean 0 and covariance function

$$(3.1) \qquad \sigma(t_1,t_2) = t_1(1-t_2)\cdot q(t_1)\cdot q(t_2)\ , \quad 0 \le t_1 \le t_2 \le 1$$

provided that $q(\cdot)$ is continuous and finite (see for instance Cox and Hinkley (1974), Appendix 2).

There is no real simplification obtained in trying to carry out the integration as in (1.3). But we can still estimate the variances and covariances of the $\hat{R}(\cdot)$ sequence by using an estimator of (3.1) in formula (5.1.1.2). This last formula is given on $[-\pi,\pi]$ so care has to be taken in reformulating (3.1) :

$$\sigma(x,y) = \left(\frac{x+\pi}{2\pi}\right)\left(\frac{\pi-y}{2\pi}\right)\cdot q\left(\frac{x+\pi}{2\pi}\right)\cdot q\left(\frac{y+\pi}{2\pi}\right) \ .$$

6.3 Conclusion

We have started Part II with some questions carried over from our empirical experience. Can we now provide some answers?

For example, we have uncovered the "odd-even" phenomenon, but we cannot explain it as such. It does not appear very markedly in estimation problems, probably because the estimators we start with are very crude.

Symmetrization works very well in the exponential case, but not so well for the chi-square. One reason could be the effect it has on the characteristic function.

For an exponential, the characteristic function goes to zero as $1/v$, but its real part, which is the characteristic function of the symmetrized exponential (the Laplace distribution), goes to zero as $1/v^2$.

For the chi-square with 4 degrees of freedom, the characteristic function goes to zero as $1/v^2$ and so does its real part. Thus there is no real gain in symmetrizing in such situations.

We have accumulated quite a lot of evidence as to transformations of data with regard to density estimation. The three approaches we used are not exactly equivalent. In Section 6.1, we underlined the assumptions behind each and, in Part I, we stressed the relations between the approaches and qualitative aspects of the data.

The best conclusion in practice would be to use the different approaches on the same data. So far it seems that the quantile approach is quite robust. We illustrate this statement by pictures of the exponential and the normal obtained via the quantile approach (see Fig. 1 and 2). In Figure 1, we have averaged the exponential approximators of orders 8 and 9. In Figure 2, we have the normal approximator of order 9.

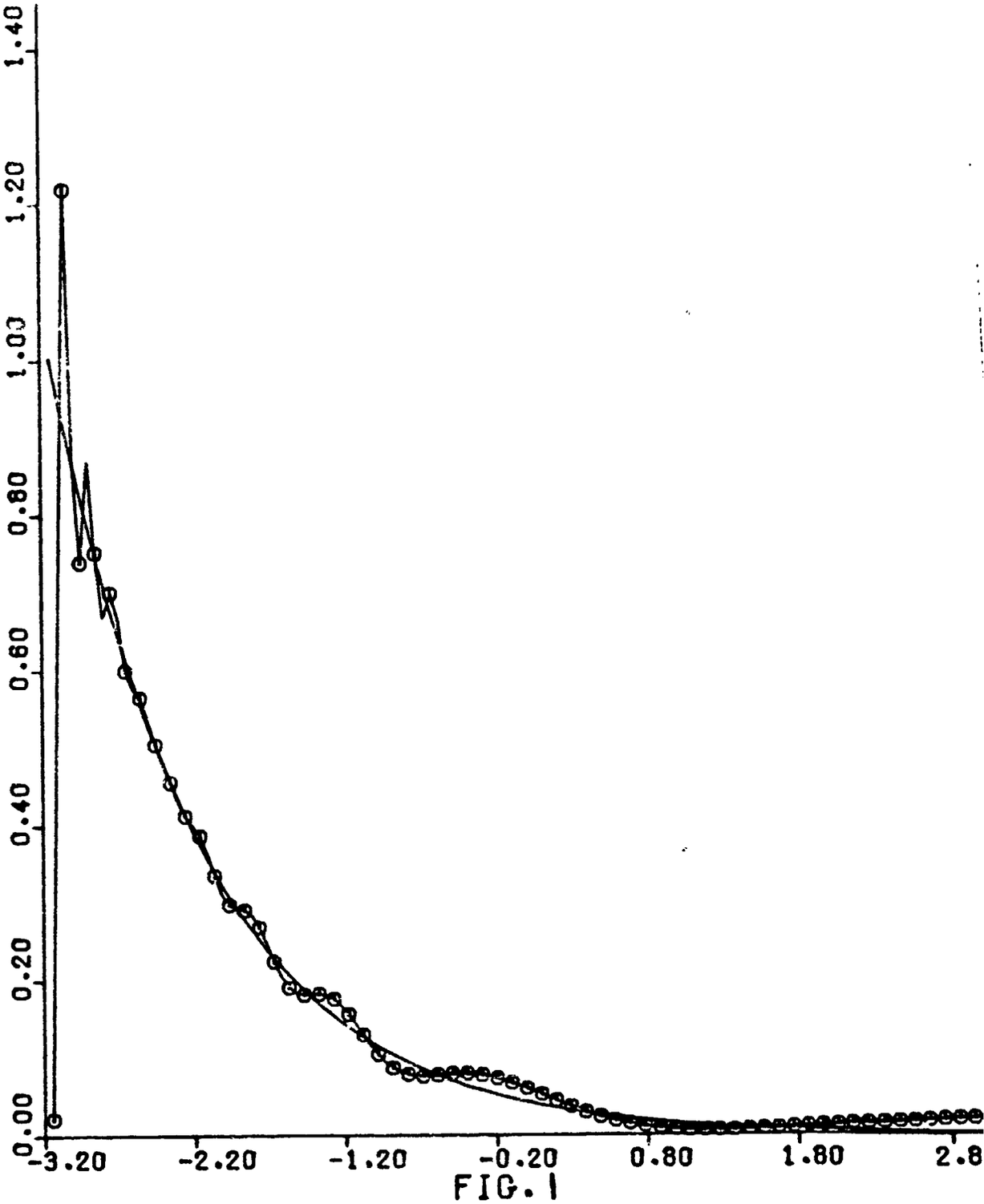


FIG. 1

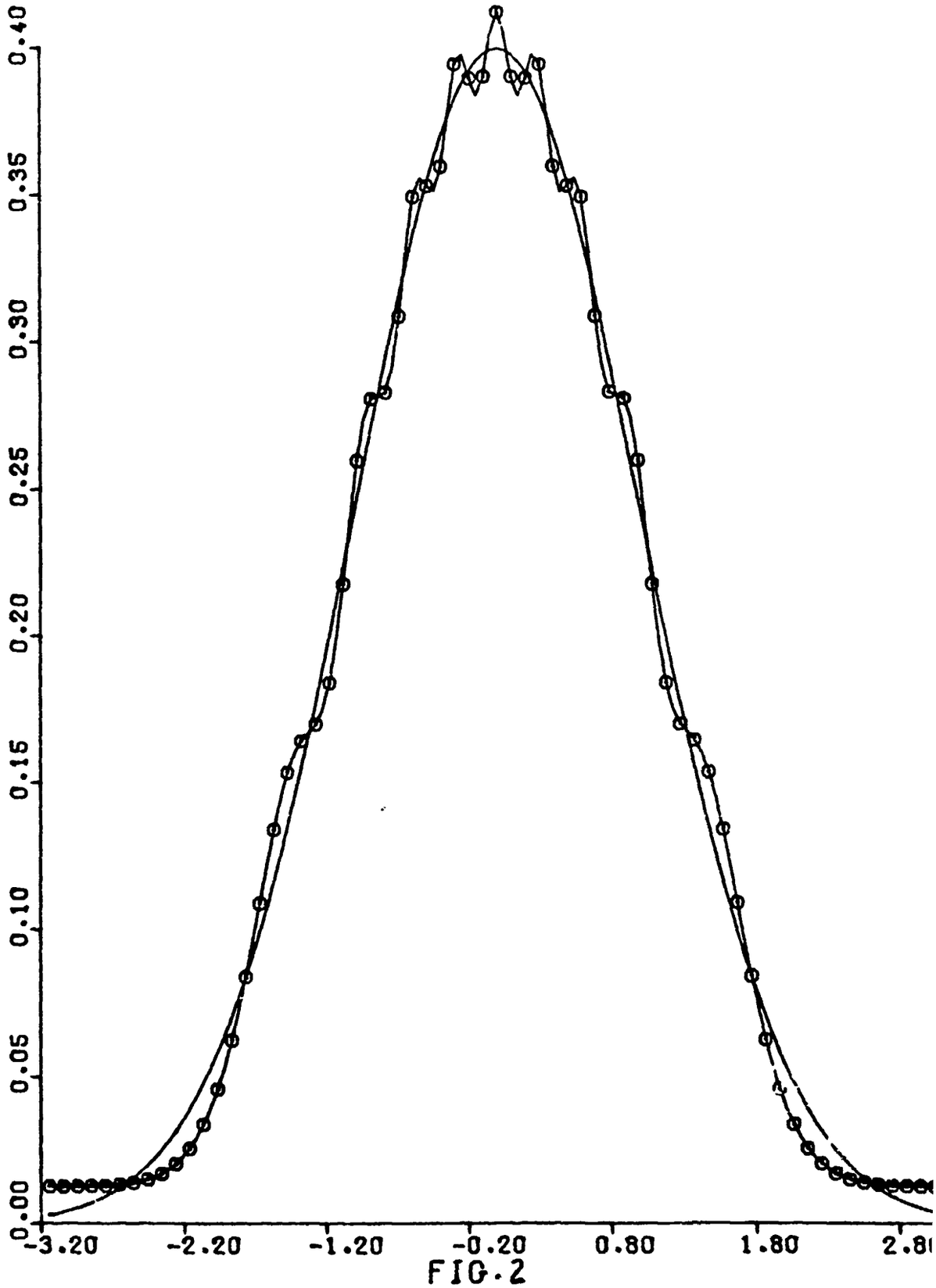


FIG.2

BIBLIOGRAPHY

Andrews, D. F., "Plots of high-dimensional data," Biometrics, 28 (1972), 125-136.

Anscombe, F. J., "Graphs in statistical analysis," American Statistician, 27 (1973), 17-21.

Berk, K. N., "Consistent autoregressive spectral estimates," Annals of Statistics, 2 (1974), 489-503.

Bliss, C. I., Statistics in Biology, vol. 1, 1967, McGraw-Hill, New York.

Boneva, L. I., Kendall, D., and Stefanov, I., "Spline transformations: three new diagnostic aids for the statistical data-analyst," Journal of the Royal Statistical Society Series B, 33 (1971), 1-37.

Chernoff, H., "Using faces to represent points in k-dimensional space graphically," Journal of the American Statistical Association, 68 (1973), 361-369.

Cleveland, W. S. and Kleiner, B., "A graphical technique for enhancing scatterplots with moving statistics," Technometrics, 17 (1975) 447-454.

Cline, A. K., "Scalar — and planar-valued curve fitting using splines under tension," Communications of the Association for Computing Machinery, 17 (1974), 218-220.

Cox, D. R., and Hinkley, D. V., Theoretical Statistics, 1974, Chapman and Hall, London.

Davies, R. B., "Asymptotic inference in stationary Gaussian time-series," Advances in Applied Probability, 5 (1973), 469-497.

Doob, J. L., Stochastic Processes, 1953, Wiley, New York.

Feder, P. I., "Graphical tools in statistical data analysis—tools for extracting information from data," Technometrics, 16 (1974) 287-299.

Feller, W., An Introduction to Probability Theory and its Applications, Vol. 2, 1966, Wiley, New York.

Geronimus, L. Y., Orthogonal Polynomials, 1961, Consultants Bureau, New York.

Grenander, U., and Szegö, G., Toeplitz Forms and Their Applications, 1958, University of California Press, Berkeley.

Ibragimov, I. A., "On the asymptotic behavior of the prediction error," Theory of Probability and its Applications, IX (1964), 627-633.

Kromer, R. E., "Asymptotic properties of the autoregressive spectral estimator," Ph.D. dissertation, Department of Statistics, Stanford University, 1969.

Kronmal, R., and Tarter, M., "The estimation of probability densities and cumulatives by Fourier series methods," Journal of the American Statistical Association, 69 (1968), 925-952.

Loftsgaarden, D. O., and Quesenberry, C. P., "A non-parametric estimate of a multivariate density function," Annals of Mathematical Statistics, 38 (1965), 1261-1265.

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A., "The time intervals between industrial accidents," Biometrika, 39 (1952), 168-180.

Mann, H. B. and Wald, A., "On stochastic limit and order relationships," Annals of Mathematical Statistics, 14 (1943), 217-227.

Pagano, M., "When is an autoregressive scheme stationary?," Communications in Statistics, 1 (1973), 533-545.

Parzen, E., "On estimation of a probability density function and mode," Annals of Mathematical Statistics, 33 (1962), 1065-1076.

Rosenblatt, M., "Remarks on some non-parametric estimates of a density function," Annals of Mathematical Statistics, 27 (1956), 832-837.

Sillitto, G. P., "Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample," Biometrika, 56 (1969), 641-650.

Sillitto, G. P., "Estimation of the median and other percentiles of an unknown continuous univariate population," Skandinavisk Aktuarietidskrift, 54 (1971), 90-96.

Thaler, H., "Non-parametric probability density estimation and the empirical characteristic function," Ph.D. dissertation, Dept. of Statistics, State University of NY at Buffalo, 1974.

Tukey, J. W., Exploratory Data Analysis, preliminary edition 1970, Addison-Wesley, Reading.

Wahba, G., "A polynomial algorithm for density estimation," Annals of Mathematical Statistics, 42 (1971), 1870-1886.

Watson, G. S., "Density estimation by orthogonal series," Annals of Mathematical Statistics, 40 (1969), 1496-1498.

Watson, G. S. and Leadbetter, M. R., "Hazard analysis II," Sankhyā Series A, 26 (1964), 101-116.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 45
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): The Autoregressive Method: A Method of Approximating and Estimating Positive Functions.
5. TYPE OF REPORT & PERIOD COVERED: (9) Technical rept.,
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Jean-Pierre Carmichael
8. CONTRACT OR GRANT NUMBER(s): (15) N00014-72-C-0508

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Statistical Laboratory
State University of New York at Buffalo
Amherst, New York 14226

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (16) NR-042-234

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistics and Probability Program Code 436
Arlington, Virginia 22217

12. REPORT DATE: (11) August 1976
13. NUMBER OF PAGES: 205 (12) 206p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (14) TR-45

15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Autoregressive schemes
Density estimation
Orthogonal Polynomial
Hazard functions
Quantile functions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~We consider~~ the Fourier transform ~~of~~ a positive function $f(\cdot)$ (or its sample Fourier transform) as a possibly complex covariance function of a hypothetical stationary complex-valued time series. We model this time series by an autoregressive process of order p whose spectral density approximates (or estimates) the function $f(\cdot)$.

We show the equivalence of this interpretation with the theory of orthogonal polynomials on the unit circle; we study the consistency of the autoregressive estimator as p increases with the sample size.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

409511